# ECG based Prediction Model for Cardiac-Related Diseases using Machine Learning Techniques

**Igor Alexandre Almeida Matias**

Dissertação para obtenção do Grau de Mestre em
**Engenharia Informática**
(2º ciclo de estudos)

Orientador: Prof. Doutor Nuno Manuel Garcia dos Santos
(Universidade da Beira Interior, Covilhã, Portugal)
Coorientador: Prof. Doutor Miguel Castelo-Branco Craveiro de Sousa
(Universidade da Beira Interior, Covilhã, Portugal)
Coorientador: Prof. Doutor Eftim Zdravevski
(Saints Cyril and Methodius University in Skopje, Skopke, North Macedonia)

**junho de 2020**

# Dedications

*“O que fizermos para nós mesmos morrerá connosco. O que fizermos para os outros e para o mundo permanece e é imortal.”*

*“What we do for ourselves dies with us. What we do for others and the world remains and is immortal.”*

Albert Pike

**Português**

Dedicada aos meus Pais, à minha Avó, à minha Irmã e ao meu Cunhado, por me apoiarem em todos os momentos, quer académicos quer pessoais, por sempre terem confiado em mim e por acreditarem no meu sucesso. A vocês devo o que hoje sou, como o sou, e a razão de o ser. Sem vocês isto não seria de todo possível.

**English**

Dedicated to my Parents, my Grandmother, my Sister and my Brother-in-law, for supporting me all the time, whether academic or personal, for always trusting me and for believing in my success. I owe you how, why and who I am today. Without you, this would not be possible at all.

# Acknowledgements

*"Excelência não é uma habilidade, é uma atitude."*

*"Excellence is not a skill, it's an attitude."*

Ralph Marston

**Português**

Neste segundo degrau de uma longa jornada, agradeço o seu completar aos três grandes elementos que o tornaram possível.

De cima, apoio-me sob o pilar direito que a minha família mais próxima me forneceu. Mãe, Pai, Avó, Irmã e Cunhado, obrigado! Obrigado pela confiança, pela paciência, pela inspiração, pela orientação, e por todos os pequenos gestos e palavras de que jamais abdicaria.

De baixo, pelo pilar esquerdo que os meus amigos mais chegados me ajudaram a construir. Gabriel, Paulo, Leonice, Henriques, Virginie e Sandeep, obrigado! Obrigado por me terem ajudado a levantar quando a motivação faltava, e, não menos importante, por me proporcionarem momentos informais que me fizeram enfrentar com maior destreza os obstáculos ao longo da jornada.

De frente, guio-me pela luz, pela sabedoria, pela exatidão e pela entreajuda que obtive em todos os passos deste jogo de xadrez monocromático que iniciei há sensivelmente seis anos. Nuno Garcia, Nuno Pombo, Vladimir Tomberg, David Lamas e João Luís, obrigado! Obrigado por permitirem o meu crescer, por me proporcionarem as experiências de vida que proporcionaram, e por me ajudarem a ter o gosto pela investigação e pelo mundo que hoje tenho. De igual modo agradeço ao Professor Miguel Castelo-Branco, ao Professor Eftim Zdravevski e à Professora Ana Gouveia, por fazerem parte e por contribuírem para esta etapa da minha vida que agora se encerra.

Por último, mas não por isso menos essencial, agradeço tudo o que me foi possível até hoje, da forma que o foi, e pela razão que o foi, pelos desafios que me fizeram crescer, pelas experiências que me fizeram pensar, e pelos acasos que me proporcionaram. Portugal e Guiné-Bissau, obrigado!

**English**

On this second step of a long journey, I thank the three great elements that made it possible.

From above, I lean on the right pillar that my closest family has provided me. Mother, Father, Grandmother, Sister and Brother-in-law, thank you! Thank you for your confidence, for your patience, for your inspiration, for your guidance, and for all the little gestures and words I would never give up of.

From below, through the left pillar that my closest friends helped me to build. Gabriel, Paulo, Leonice, Henriques, Virginie and Sandeep, thank you! Thank you for helping me to get up when my motivation was lacking, and, not least, for providing me with informal moments that made me face obstacles along the journey with greater skill.

From the front, I am guided by the light, the wisdom, the accuracy and the mutual assistance that I obtained in all the steps of this monochrome chess game that I started about six years ago. Nuno Garcia, Nuno Pombo, Vladimir Tomberg, David Lamas and João Luís, thank you! Thank you for allowing me to grow, for providing me with the life experiences that you have provided, and for helping me have a taste for research and the world I have today. In the same way, I would like to thank Professor Miguel Castelo-Branco, Professor Eftim Zdravevski and Professor Ana Gouveia, for being part of and contributing to this stage of my life that is now ending.

Last but not least, I am grateful for everything that has been possible to date, in the way that it was, and for the reason that it was, for the challenges that made me grow, for the experiences that made me think, and for the chance that they provided me. Portugal and Guinea-Bissau, thank you!

Dissertation prepared at the Ambient Assisted Living Computing and Telecommunications Laboratory (ALLab), *Instituto de Telecomunicações*, *Universidade da Beira Interior*, and submitted to the *Universidade da Beira Interior* for defense in a public examination session.

The research has been partially funded by the FCT/MCTES through national funds, and when applicable, co-funded EU funds under the project UIDB/EEA/50008/2020 and Operação Centro- 01-0145-FEDER-000019 – C4 – Centro de Competências em Cloud Computing, co-financed by the Programa Operacional Regional do Centro (CENTRO 2020), through the Sistema de Apoio à Investigação Científica e Tecnológica – Programas Integrados de IC&DT. I would also like to acknowledge the contribution of the COST Action IC1303: AAPELE—Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226; SHELD-ON—Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology).

# Resumo

Esta dissertação descreve a construção de modelos preditivos de condições de saúde através de aplicação de métodos de Inteligência Artificial. O trabalho é assim focado na predição, a curto e longo prazo, de condições de Fibrilhação Auricular através da análise de exames de Eletrocardiografia, com a utilização de diversas técnicas de redução de ruído e de interferência, bem como a sua representação através de espectrogramas e sua aplicação em modelos de Inteligência Artificial, concretamente de Aprendizagem Profunda (*Deep Learning* na língua inglesa). Os processos de treino e teste dos modelos obtidos recorreram a uma base de dados publicamente disponível. Nas suas duas abordagens, foram obtidos algoritmos preditivos com uma precisão de 96.73% para uma predição de curto horizonte e 96.52% para longo horizonte de predição de Fibrilhação Auricular. Os objetivos principais da presente dissertação são assim o estudo de trabalhos já realizados na área durante a última década, apresentar uma nova metodologia de predição da condição apresentada, bem como apresentar e discutir os seus resultados, incluindo sugestões de melhoria para futuro desenvolvimento.

# Palavras-chave

Fibrilhação Auricular; Eletrocardiograma; Algoritmos Preditivos na Saúde; Aprendizagem Profunda; Rede Neuronal Convolucional; Espectrograma; Inteligência Artificial

# Abstract

This dissertation presents research on the construction of predictive models for health conditions through the application of Artificial Intelligence methods. The work is thus focused on the prediction, in the short and long term, of Atrial Fibrillation conditions through the analysis of Electrocardiography exams, with the use of several techniques to reduce noise and interference, as well as their representation through spectrograms and their application in Artificial Intelligence models, specifically Deep Learning. The training and testing processes of the models made use of a publicly available database. In its two approaches, predictive algorithms were obtained with an accuracy of 96.73% for a short horizon prediction and 96.52% for long Atrial Fibrillation prediction horizon. The main objectives of this dissertation are thus the study of works already carried out in the area during the last decade, to present a new methodology of prediction of the presented condition, as well as to present and discuss its results, including suggestions for improvement for future development.



# Keywords

Atrial Fibrillation; Electrocardiogram; Predictive Algorithms in Healthcare; Deep Learning; Convolutional Neuronal Network; Spectrogram; Artificial Intelligence

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| acc | Accuracy |
| AF | Atrial Fibrillation |
| AFPD | Atrial Fibrillation Prediction Database |
| AI | Artificial Intelligence |
| ARFC | Arrhythmia Fuzzy Hybrid Classifier |
| CAF | Chronic Atrial Fibrillation |
| CBC | Complete Blood Count |
| CMP | Complete Metabolic Profile |
| CNN | Convolutional Neural Network |
| CR | Chest Radiography |
| CVD | Cardiovascular Diseases |
| DFT | Discrete Fourier Transform |
| ECG | Electrocardiogram/Electrocardiography |
| ECG-LVH | Electrocardiographic Criteria for Left Ventricular Hypertrophy |
| Echo | Echocardiography |
| FFT | Fast Fourier Transform |
| FFT-HF | High-frequency component of Fast Fourier Transforms |
| FFT-LF | Low-frequency component of Fast Fourier Transforms |
| FN | False Negative |
| FP | False Positive |
| HF | High-frequency band power |
| HRV | Heart Rate Variability |
| ICU | Intensive Care Unit |
| LAA | Left Atrial Appendance |
| LF | Low-frequency band power |
| LP | Long Prediction |
| LPSA | Long Prediction Simple Approach |
| ML | Machine Learning |
| NN50 | Number of adjacent RR intervals differing by more than 50 milliseconds |
| PAC | Premature Atrial Complexes |
| PAF | Paroxysmal Atrial Fibrillation |
| PNN50 | Sum of NN50 divided by the total number of all RR intervals |

| | |
|---|---|
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| PTF | P Wave Terminal Force |
| PWA | P Wave Axis |
| RMSSD | Square root of the mean of the squares of differences between adjacent RR intervals |
| ROI-WCOB | Weighted centre of the bispectrum |
| SD | Standard Deviation |
| SDANN | Standard Deviation of Average of all NN interval for all 5-minute periods of the entire recording |
| SDRR | Standard Deviation of RR intervals |
| SDSD | Standard deviation of differences between adjacent RR intervals |
| SP | Short Prediction |
| SPHA | Short Prediction Hybrid Approach |
| SPSA | Short Prediction Simple Approach |
| SPWVD | Smoothed Pseudo Winger Ville distribution |
| STFT | Short-Time Fourier Transform |
| SVM | Support Vector Machine |
| THM | Thyroid-stimulating Hormone Measurement |
| TN | True Negative |
| TP | True Positive |
| UBI | Universidade da Beira Interior |
| val_acc | Validation Accuracy |
| val_loss | Validation Loss |
| VLF | Very Low-Frequency band power |
| WHO | World Health Organization |

# Chapter 1

# 1. Introduction

## 1.1. Motivation

The World Health Organization (WHO) states that Cardiovascular Diseases (CVDs) are associated with heart and blood vessels-based disorders. CVDs may include hypertension, heart attack, cerebrovascular disease, heart failure, rheumatic congenital heart disease, and cardiomyopathies, among others [1].

CVDs were the most significant cause of death worldwide in 2016 [1], adding to approximately 27.00% of all the estimated deaths in the world [2] and about 45.00% in Europe by 2017 [3]. Solely in the European Union, in the year of 2006, CVDs were estimated to have cost €104.00 thousand million in healthcare [4].

Among the CVDs, Atrial Fibrillation (AF) is an arrhythmia, which is characterized by irregular heartbeats, that can lead to blood clots, heart failure, stroke, and other heart-related complications including death, and is commonly underdiagnosed [5], [6]. It can assume four different types, Paroxysmal AF, Persistent AF, Long-standing Persistent AF, and Chronic/Permanent AF. According to [7], only the severest Long-standing Persistent and Permanent can be easily detected with an Electrocardiogram (ECG) exam, the other types being harder to identify due to the irregularity of its manifestations

This irregularity makes it very hard to detect both less severe types of AF, due to the high probability of these patients not presenting any symptoms during an occasional ECG. To avoid this low efficiency of detection of AF, predictive models were being developed, allowing the diagnosis of a patient's AF state only based in a short ECG signal, avoiding extra-long and intrusive devices and methodologies.

Typically, AF can be diagnosed using pulse palpation, ECG prolonged monitoring with portable devices, or by using pacemakers and other types of implanted devices, however, the detection of AF is commonly performed by analysing the signal collected from an ECG, a non-invasive and painless exam with quick results, typically outputting several charts resulting from a 12 lead collection setup [8].

Nowadays, portable devices such as smartwatches, smart fitness bands or portable medical signal collectors have a crucial role in evolving the way we diagnose several health disorders before they step into a high-risk medical field. It is due to its ease on the recording, for example, ECG and pulse signals [9]–[12], being always present with the patient itself, being thus able to collect data from several moments of the day, within different activity and emotional states.

Some of these devices, despite using a smaller number of ECG leads, sometimes 3 or 2, have been proved to be as efficient as Hospital grade ECG equipment, as tested in [13].

However, in the last years, there were developed several new methods to detect and to predict the existence of the different types of AF. These new approaches all require powerful algorithms combined with innovative sensors, applying several different types of Artificial Intelligence (AI).

There is a multitude of benefits from integrating AI into healthcare, including automation tasks and analysing big patient data sets to deliver better healthcare faster, and at a lower cost [14]. The usage of AI into healthcare, and consequently AF detection and prediction, does allow the analysis of bigger datasets, with the faster result, easing the workload of healthcare professionals, making possibly automated and real-time diagnosis, anytime and anywhere.

Thus, the main goal of the presented work is to provide a new approach on the diagnosis of AF as well on the prediction of AF episodes, making possible to better prepare, or even revert, an AF episode start, this way preventing many possible severe health conditions. This work is presented as a Machine Learning (ML) method applied to ECG exam signals.

## 1.2. Main Contributions

This master's dissertation introduces several contributions to different aspects of visual recognition. However, the work was focused on the classification of signals before an Atrial Fibrillation (AF) episode, to develop an algorithm able to predict the onset of a Paroxysmal Atrial Fibrillation (PAF) episode. The main contributions of this dissertation are:

- In the third chapter, the state of the art details how AI methods achieve good accuracy and can be applied on the prediction of AF episodes, describing and clarifying the architectures, and its limits, applied by work performed worldwide since the year of 2009;
- The fourth chapter describes an innovative approach to applying DL methods to Electrocardiographic data converted into spectrogram images, applying noise reduction and data optimization techniques. This chapter contains all the detail needed to replicate this study and improve its work;
- In chapter five are presented the results of the work described in the previous one, resulting in three final trained, validated and tested DL models for the prediction of PAF episodes in short and long periods, using simple and hybrid approaches combined with the innovative techniques applied and described in chapter four;
- Finally, chapter six contains this study's final remarks and conclusions, as well as some studied techniques to overcome the limited capacity of the CNN model applied by this study whit the used data;

- Additionally, this work also resulted in the production of two scientific papers, already submitted to peer-reviewed journals, focusing the systematic literature review and the results achieved.

## 1.3. Organization of the Document

This dissertation is organized as follows:

1. The first chapter – Introduction – presents the motivation for this work and its main contributions;
2. The second chapter – Atrial Fibrillation – provides an introduction and context on the health condition in which this work is focused;
3. The third chapter – State of the Art – contains a study of the work performed in a similar context as this, spanning the last ten years;
4. The fourth chapter – Methodologies – presents an introduction of each one of the three major technologies/techniques used, and describes each part of the work, justifying and relating each one of the steps taken since the design until the results phases;
5. The fifth chapter – Results – provides the results achieved by this work and performs a comparison with the previously worked approaches mentioned by the third chapter;
6. The sixth chapter – Conclusions – contains the discussion and conclusion of the results and applied methods, as well as possible future improvements.

# Chapter 2

# 2. Atrial Fibrillation

## 2.1. Introduction

Atrial Fibrillation (AF) is the most common cardiac arrhythmia, which, according to [15], affects about 1.00 per cent of patients below 60 years old and about 8.00 per cent patients above 80 years old. As stated in [16]–[20], it is relatively uncommon for patients under 50 years old, but occurs in 2.00 to 4.00 per cent of the population after 65 years old, doubling the risk with each advancing decade of life.

AF is defined as a supraventricular tachyarrhythmia characterised by uncoordinated atrial activation, consequently deteriorating mechanical functions of the hearts' atrium chamber [21], thus being a source of significant morbidity and mortality because it impairs cardiac function, increasing the risk of stroke.

There are different types of AF, which have different prognoses, morbidity rates, mortality rates and treatment options [21], in which Paroxysmal Atrial Fibrillation (PAF) refers to episodes of intermittent AF that terminate spontaneously, and Chronic Atrial Fibrillation (CAF) to continuous episodes that cannot be converted back to normal sinus rhythm with no medical help [22].

However, as shown in [23], one year after the initial diagnosis of PAF, a patient state progress to CAF with a probability of 8.60%, and this value increases to 24.70% five years after.

This chapter presents a short introduction on the AF and PAF conditions, regarding its effects, symptoms and commonly used diagnostic techniques.

The main contributions of this chapter are:

1) present a theoretical and basic explanation of the AF health condition (2.1);

2) express its most common effects on a person health (2.2);

3) present the frequently shown symptoms that help on the diagnosis (2.3);

4) introduce the used techniques for its diagnosis, as well as briefly explaining each one (2.4).

## 2.2. Effects

A diagnosed PAF patient has a high probability of developing blood clots, heart failure, stroke, valvular disease, and hypertension [5], [6], [23], directly related to the loss of coordinated atrial contraction caused by an AF episode, in which occurs changes in depolarization patterns of the atrium, and apoptosis of the cells [24], [25]. A restoration of a sinus rhythm becomes more difficult as the time of heart fibrillation increases.

As clinical implications, AF can lead to rapid ventricular response, decreased diastolic filling or blood stasis and atrial clot formation, which, in the end, represent an increased morbidity and mortality rates. It's most common clinical implications are shown in Fig. 1, as stated by [22].

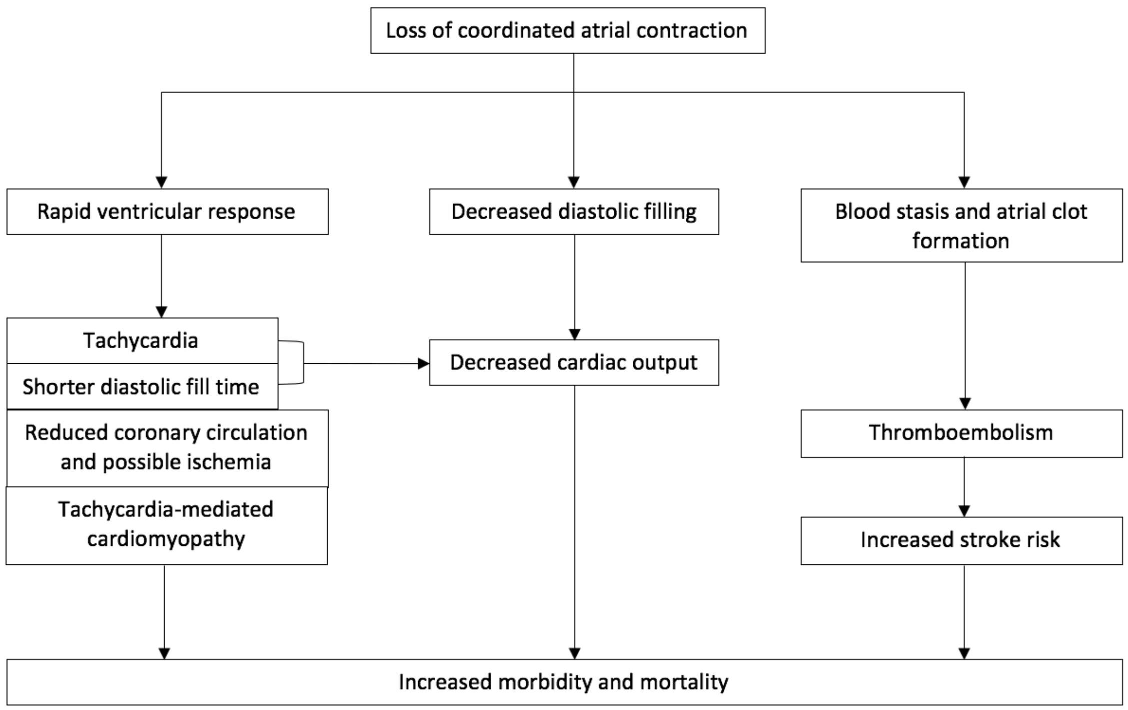



*Figure 1 - Flowchart of clinical implications of AF* [22].

A rapid ventricular response is characterised as "a heart rhythm disorder (arrhythmia) caused by abnormal electrical signals in the lower chambers of the heart (ventricles)" [26], and is related to a shorter diastolic fill time, which can lead to sudden death, according to [27], thus decreasing the cardiac output, that is, the cardiac performance. Cardiac Ischemia is the deficient oxygenation of myocardial cells, that can be caused by obstruction of the coronary flow or by other reason [28], while Cardiomyopathy is associated to a disease that makes the heart muscle hard to pump blood to the body, leading to heart failure [29]. Blood stasis is associated with atrial clot formation, which leads to Thromboembolism, that is, migration of a blood clot originated in the Left Atrial Appendance (LAA) [30], leading to an increased stroke risk.

## 2.3. Symptoms

AF presents a wide spectrum of symptoms and signs, some patients are asymptomatic, and only the two severest types can be easily detected [7].

The manifestations of an AF patient range from asymptomatic to stroke, heart failure and cardiovascular collapse, the most common being palpitations, dyspnea, fatigue, lightheadedness, and chest pain [22]. However, because symptoms are nonspecific, that is, not exclusive or associated to a few specific health conditions, and not always detectable, they cannot be used to diagnose and determine the onset of an AF episode, that is, the onset of a PAF episode.

The commonly used tools for the diagnosis of AF are presented in the next section.

## 2.4. Diagnosis

According to [31], there are four main approaches to diagnose AF, specifically the silent AF, that is, the asymptomatic AF cases. This is possible by screening for AF in the community, by a prolonged monitoring for PAF, by monitoring patients with pacemakers and implanted devices, or even with detection in stroke survivors.

Screening for AF in the community is used to prevent very common undiagnosed AF in older populations and patient with heart failure [32], using either short-term ECG or pulse palpation (followed by ECG in those with an irregular pulse).

The prolonged monitoring for PAF procedures include repeated daily ECG recordings, which increased the detection of PAF in an unselected Swedish population aged above 75 years [33], [34], can be performed using several patient-operated devices [35], [36] and extended continuous ECG monitoring patch recorders [37]. Also, new technologies such as smartphone cases with ECG electrodes, smartwatches and blood pressure machines with AF detection algorithms, can be used [38].

Patients with atrial high rate episodes can be identified by using implanted pacemakers or defibrillations with atrial lead, which allow a continuous monitoring of atrial rhythm, thus making it possible to possibly detect PAF episodes by further assessment of stroke risk factors, including ECG monitoring. Depending on the risk profile of the studied population, atrial high rate episodes are detected in 10.00 to 15.00% of pacemaker patients [39].

In stroke survivors, AF can be detected with a prolonged ECG monitoring, by which the study [40] detected AF in 24.00% of stroke survivors, with variations depending on the timing, duration and method of monitoring. AF detection is more likely in patients with cryptogenic stroke implanted with loop recorders or who was ECG monitored for several weeks [5], [6], [41].

Cryptogenic stroke is a type of stroke that could not be linked to a specific cause after extensive investigations [42].

## 2.5. Evaluation

After being detected, there is a need to find the clinical reason for AF appearance, that is, to discover if the AF condition was induced by another health condition or not.

Thus, the standard and initial evaluation of AF is made, according to [22], with six types of tests, which are Chest Radiography (CR), Complete Blood Count (CBC), Complete Metabolic Profile (CMP), Echocardiography (Echo), Electrocardiography (ECG) and Thyroid-stimulating Hormone Measurement (THM). The first one identifies a possible pulmonary disease, such as pneumonia, vascular congestion or chronic obstructive pulmonary disease. CBC procedure is responsible to identify comorbid conditions, for example, anaemia or infections, while CMP is capable of identifying electrolyte abnormalities that may cause or exacerbate atrial fibrillation. Both Echo and ECG exams are related to the patient's coronary system, where the first one assesses heart size and shapes, chamber sizes and pressures, valve structure and function, detect the presence of pericardial effusion, wall motion abnormalities, and systolic and diastolic dysfunction. ECG diagnoses AF and identify other arrhythmia or cardia conditions, such as left ventricular hypertrophy or ischemia, using the electrical signals collected from the patient’s body, which are related to the heart muscle activity. At last, THM is capable of identifying hyperthyroidism conditions.

# Chapter 3

# 3. State of the Art

## 3.1. Introduction

This chapter presents a systematic literature review on ECG-based models for AF Prediction using AI techniques covering the last ten years. At the time of this review, there were not found any report that covers this topic.

Its contents have already been submitted to the Artificial Intelligence in Medicine scientific journal.

Therefore, the selected studies reviewed here present the most recent work in this field, and the main contributions of this chapter are:

1) present a discussion on how the prediction of AF has been and is currently addressed;

2) indicate what databases, features, pre-processing and predictive algorithms have been and are presently used in these systems;

3) a benchmark to conclude which achieved models performs better.

The remainder of this chapter is organized as follows: Section Methods presents a description of the method that was designed for eligibility selection and extraction of information. Section Results includes the results of the search by displaying the selected studies and their features in summary tables. The discussion and the answer to the research questions are presented in Section Discussion. Finally, Section Conclusion of this chapter shows the highlights and limitations of this state of the art study.

## 3.2. Methods

This systematic literature review was conducted informed by recommendations from the Cochrane Handbook for Systematic Reviews of Interventions, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [43]–[45], and based on the guidelines from [46].

This section explains in detail the methodology used for conducting this review.

### 3.2.1. Search Strategy

The databases Web of Science[1], Scopus[2], ACM Digital Library[3], IEEE Xplore[4], PubMed[5] and Science Direct[6] were used to search for relevant peer-reviewed publications from January 1, 2009, 00:00 to December 13, 2019, 04:22.

The first two used databases are interdisciplinary databases. ACM Digital Library is, according to [47], the number one database related to academic databases for computer science and IEEE Xplore was chosen due to its high number of articles from the field of computer science. Finally, PubMed was used due to its content regarding research in biomedicine and Science Direct because of its high number of articles from thousands of books and journals.

There were searched titles and abstracts using the keywords presented below. The list of references from the selected articles was manually screened for the inclusion of additional relevant articles.

The keywords used in all the databases were:

> *("machine learning" OR "artificial intelligence") AND ("ECG" OR electrocardio*) AND ("Atrial Fibrillation" OR "AF" OR "arrhythmia") AND ("prediction" OR "prognosis" OR "foresee").*

### 3.2.2. Study Selection

There were screened the titles and abstracts of all identified publications for eligibility, using the web application Rayyan QCRI [48].

The inclusion criteria were broadly defined to increase the sensitivity of the search. The aim was to identify the articles that applied any AI method on ECG signals for prediction of AF on patients with no previous clinical conditionings.

Additional inclusion/exclusion criteria are summarized in Table 1.

According to the inclusion and exclusion criteria presented in Table 1, all the articles not excluded after its analysis had its full texts reviewed for eligibility.

[1] S. C. Collection, "Web of Science [v.5.14] - Web of Science Core Collection引用レポート," pp. 8–9
[2] "Scopus - Document search | Signed in." [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 15-Jan-2020]
[3] C. The et al., "ACM Digital Library," 1985
[4] "IEEE Xplore Digital Library." [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: 15-Jan-2020]
[5] "Home - PubMed - NCBI." [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/. [Accessed: 15-Jan-2020]
[6] "ScienceDirect.com | Science, health and medical journals, full-text articles and books." [Online]. Available: https://www.sciencedirect.com/. [Accessed: 15-Jan-2020]

*Table 1 - Inclusion and exclusion criteria used in the review, as in* [49].

| Type | Inclusion | Exclusion |
|---|---|---|
| Date | All | None |
| Exposure of interest | All | None |
| Geographic location of study | All | None |
| Language | English | Any other language |
| Participants | With no recent surgical procedures or drugs effects during the ECG collection | With any recent surgical procedure or ingestion or drugs effects during the ECG collection |
| Peer review | Journal and Conference | All others |
| Reported outcomes | At least one: accuracy, sensitivity, specificity, confusion matrix | All others that did not report any metric |
| Setting | All | None |
| Study design | All | None |
| Type of publication | Journal and Conference | All others |

### 3.2.3. Extraction of Study Characteristics

The extraction of information from the selected publications was based on the pre-defined categories, to collect the relevant data and to assess, analyse the model characteristics and its experimental setup:

- Study Information: defines the study citation and year of publication;
- Inputs: assess the inputs used to develop the algorithm, including used dataset and amount and age of the individuals from where the dataset was collected;
- Signal treatment: defines the usage of the ECG signals received as input, namely the features extracted from it, the duration of the signal used for training, and the tools used for the process;
- Methods: defines the methods/algorithms applied to the pre-processing of the ECG signal, the prediction of AF and evaluation of the model, as well as the number of iterations, and the data split applied for training and testing;
- Performances: defines the evaluation metrics used to assess the predictions.

### 3.2.4. Research Questions

The research questions of this review were:

- (RQ1) How is the prediction problem assessed?
- (RQ2) What databases and features are used?
- (RQ3) What pre-processing algorithms are used?

- (RQ4) What predictive algorithms are used?
- (RQ5) Which are the models that perform the best?

The (RQ1) motivation was to identify the trends and possible opportunities for research topic focus.

The motivation for (RQ2) and (RQ3) was to identify new advances on features and databases and pre-processing techniques used for prediction of AF, respectively.

The motivation for (RQ4) was to identify the new predictive algorithms used to predict AF using ECG data on recent studies.

Finally, for (RQ5) motivation, it was intended to identify the models that can more accurately predict AF episodes, this way identifying trends and possible opportunities for the use of research methods.

## 3.3. Results

At the beginning of the search, it yielded 375 unique records, after the removal of duplicates.

After the review of the title and abstract and following the inclusion and exclusion criteria presented in Table 1, 293 records were excluded; 82 full-text publications were assessed for eligibility and after full-text review, of which 72 records were excluded (Fig. 2).

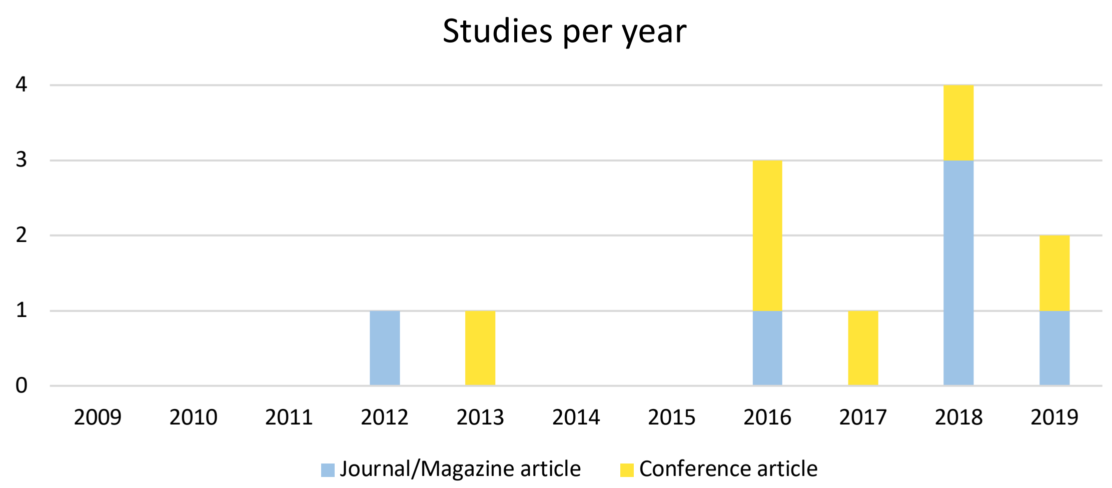



*Figure 2 - Number of studies from 2009 to 2019.*

The excluded records can be described as follows. Sixty-four studies reported research related to AF, but there was no prediction of AF during its execution. Two studies could not be fully read because the authors of this systematic literature review were not able to obtain the full articles. Two articles did not present the evaluation metrics included in the Inclusion Criteria of this search presented in Table 1. Two studies were focused on reviewing state-of-the-art related to

AF identification. One study had a publication date before 2009, and another one did the work with ECG collected from patients with surgical proceedings (prophylactic ICD-implantation).

From the remaining 10 records, reference tracking was performed, and two studies were added, totalizing 12 studies to be included for the data extraction and the qualitative synthesis stage. The flow diagram of the identification and inclusion of articles is shown in Fig. 3.

### 3.3.1. Eligibility of the Studies

Despite all the selected studies that met the inclusion and exclusion criteria, it is useful to clarify the selection of some studies.

The study [50] presents an algorithm for short term prediction of Persistent AF, but the ECG data used was collected from sheep instead of human individuals. Despite this, it was considered that this article is eligible, not so much because of the nature of the ECG signal, but mostly because of the described methodologies and algorithmic approaches the paper describes.

In the studies [51], [52] and [53] the prediction of AF was only performed between pre and post AF moments, not allowing for cases with no AF prediction. However, they were included because of the insight the papers report to this research.

Finally, in the study [53] the reported measurements with a single fold method match neither the tables nor the text of the paper. It was decided to include this last study, but only to consider the best measurements for the 10-fold method, that has valid reporting of values in the tables and the study's text.

### 3.3.2. Source of Evidence

To be able to evaluate and classify the selected studies, Table 2, Table 3, Table 4, Table 5 and Table 6 present statistical data about the publication year, the ranking of the magazine the article was published in, the type of publication, the geographic region the study covered and the number of citations the article has.

*Table 2 - Number of publications by year of publication.*

| | **Description** | **Number of studies** | **Portion of total** |
|---|---|---|---|
| Publication year | 2009 – 2016 | 5 | 41.67% |
| | 2017 – 2018 | 5 | 41.67% |
| | 2019 | 2 | 16.67% |

*Table 3 - Ranking of the article's publication place, according to* [54] *and* [55].

| | **Description** | **Number of studies** | **Portion of total** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Ranking of article's publication place, accordingly to [54] and [55] | 1st quarter Journal | 5 | 41.67% |
| | 4th quarter Journal | 1 | 8.33% |
| | H-factor = 83 Journal | 4 | 33.33% |
| | H-factor = 23 Journal | 1 | 8.33% |
| | H-factor = 700 Journal | 1 | 8.33% |
| | 0 <= H-factor <= 10 Conference Proceeding | 3 | 25.00% |
| | 11 <= H-factor <= 20 Conference Proceeding | 2 | 16.67% |
| | B1 ranked Conference | 2 | 16.67% |
| | B4 ranked Conference | 1 | 8.33% |

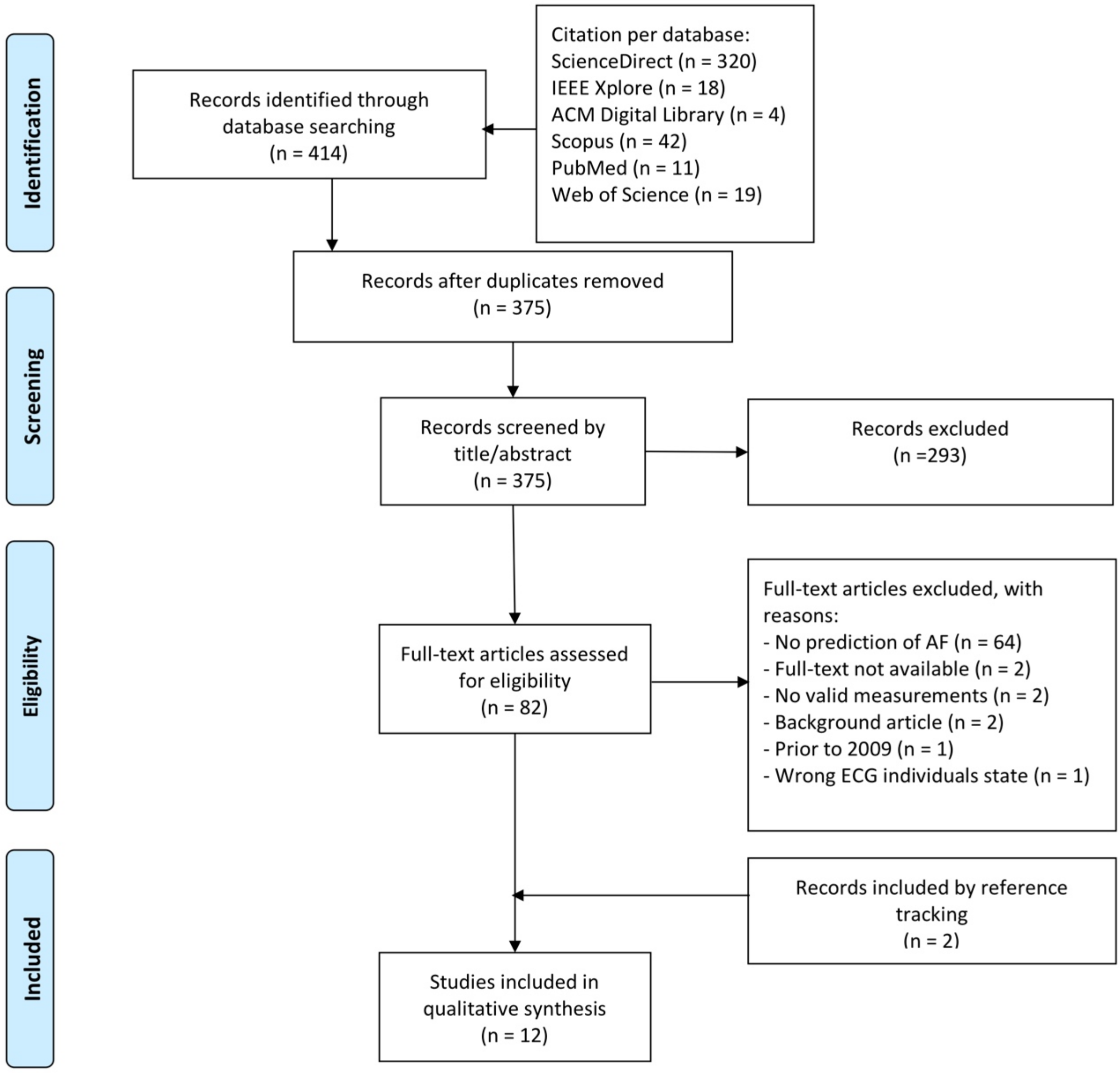



*Figure 3 - Flow diagram of identification and inclusion papers.*

*Table 4 - Number of publications by journal or conference type.*

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| Journal or Conference type | Medicine Journal | 1 | 8.33% |
| | Bioinformatics Journal | 4 | 33.33% |
| | Computer Science Journal | 1 | 8.33% |
| | IEEE Conference or sponsored by IEEE | 3 | 25.00% |

*Table 5 - Continent where the studies were conducted.*

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| The continent where it was conducted | Africa | 1 | 8.33% |
| | Asia | 6 | 50.00% |
| | Europe | 3 | 25.00% |
| | America | 2 | 16.67% |

*Table 6 - Number of citations per selected articles, according to [56].*

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| Publication citations, according to [56] | 0 | 4 | 33.33% |
| | 2 | 1 | 8.33% |
| | 4 | 1 | 8.33% |
| | 5 | 1 | 8.33% |
| | 7 | 1 | 8.33% |
| | 8 | 1 | 8.33% |
| | 12 | 1 | 8.33% |
| | 15 | 1 | 8.33% |
| | 48 | 1 | 8.33% |

### 3.3.3. Study Participants and Design

Seven studies (58.33%) were based on databases with small samples of individuals (less than 100), two studies (16.67%) with samples between 100 and 25000 individuals, one study (8.33%) with a sample of around 126000 individuals, and two studies (16.67%) did not report the sample size.

Three studies (25.00%) used a personal database of ECG records, one (8.33%) used a UCI Repository Warehouse's ([57]) database, one was based on the Mayo Clinic ([58]) ECG Laboratory's database, one used the Medical Information Mart for Intensive Care III database

([59]), five articles (41.67%) have done the research using the Atrial Fibrillation Prediction Database ([60]), and one study (8.33%) used a China Kadoorie Biobank's ([61]) database. Both [59] and [60] datasets are available at the Physionet Repository. Only [59], [60] and [61] are publicly available.

### 3.3.4. Prediction Methods

The selected articles used several different methods of AI for prediction of AF:

- Five articles ([50], [52], [53], [62], [63]) (41.67%) applied Support Vector Machine;
- Two articles ([64], [65]) (16.67%) used statistical AI methods;
- Two articles ([66], [67]) (16.67%) used Convolutional Neural Network;
- One study ([68]) (8.33%) applied its study using Arrhythmia Fuzzy Hybrid Classifier;
- Another study ([69]) (8.33%) used Markov Chain;
- Finally, the last of all articles ([51]) used the method Mixture of Experts for prediction of AF.

### 3.3.5. Data Collected from Selected Studies

During the quality synthesis process, it is essential to get as much information from the selected studies as possible. However, despite all the articles having some extra data, some of it was not comparable, the reason why they are not mentioned in the next collected data tables.

Table 7 shows the dataset used in each one of the selected studies, including the number of individuals, where the data came from, and its age, if provided.

From all the twelve selected studies, not all of them apply AI methods that do not need feature selection and extraction from the source ECG signal. Table 8 presents the number of frequency-domain, time-domain, space-domain and non-linear features extracted from each one of the studies, as well as the signal duration used as input to the AI model/method and the tools used at the collecting and pre-processing phase of the studies.

Regarding the signal duration used in each one of the selected studies, it is a noticeable difference between the minimum and maximum among all. The majority used a signal of 300 seconds (three studies), followed by 30 and 10 seconds (two studies each). Some other articles reported usage of signals with 120, 1800 and 3600 seconds length (one study each).

*Table 7 - Input information collected from studies.*

| Year | Study | Dataset used | Number of participants | Age of participants |
|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [52] | Atrial Fibrillation Prediction | - | - |

| | | | | |
|---|---|---|---|---|
| | | Database [60] | | |
| 2013 | Costin *et al.* [64] | Atrial Fibrillation Prediction Database [60] | 75 | - |
| 2016 | Kim *et al.*[66] | Own collected dataset | 1 | - |
| 2016 | Shen *et al.*[62] | China Kadoorie Biobank [61] | 24369 | - |
| 2016 | Boon *et al.*[53] | Atrial Fibrillation Prediction Database [60] | 53 | - |
| 2017 | ElMoaqet *et al.*[50] | Own collected dataset | 33 | - |
| 2018 | Rajalakshmi *et al.*[68] | UCI Repository Warehouse [57] | - | - |
| 2018 | Li *et al.*[69] | Own collected dataset | 5 | - |
| 2018 | Boon *et al.*[63] | Atrial Fibrillation Prediction Database [60] | 53 | - |
| 2018 | Ebrahimzadeh *et al.* [51] | Atrial Fibrillation Prediction Database [60] | 53 | - |
| 2019 | Attia *et al.*[67] | Mayo Clinic ECG Laboratory [58] | 126526 | >18, average 60.30 |
| 2019 | Mohamed *et al.*[65] | Medical Information Mart for Intensive Care III database [59] | 246 | - |

*Table 8 - Signal treatment information collected from studies.*

| Year | Study | Features extracted from ECG signal | Signal duration (seconds) | Tools used |
|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [52] | 4 frequency-domain | 300 | - |

| | | | | |
|---|---|---|---|---|
| | | 6 time-domain<br>4 non-linear | | |
| 2013 | Costin *et al.* [64] | 1 frequency-domain<br>1 time-domain | 300 | Pan-Tompkins algorithm [70], MATLAB 2008 [71] |
| 2016 | Kim *et al.*[66] | NR | 30 | Caffe deep learning framework [72] |
| 2016 | Shen *et al.*[62] | 1 time-domain<br>1 space-domain | 10 | - |
| 2016 | Boon *et al.*[53] | 8 frequency-domain<br>1 time-domain | 1800 | - |
| 2017 | ElMoaqet *et al.*[50] | 1 frequency-domain<br>5 time-domain<br>3 non-linear | 30 | MATLAB [71], LibSVM toolbox [73] |
| 2018 | Rajalakshmi *et al.*[68] | 5 time-domain | - | Excel[7], MATLAB 2015 [71], Rapid Miner[8] |
| 2018 | Li *et al.*[69] | NR | 120 | NR |
| 2018 | Boon *et al.*[63] | 3 frequency-domain<br>2 time-domain<br>2 non-linear | 900 | C++ [74], LibSVM library [73] |
| 2018 | Ebrahimzadeh *et al.* [51] | 4 frequency-domain<br>5 time-domain<br>8 non-linear<br>11 time-frequency | 300 | - |
| 2019 | Attia *et al.*[67] | NR | 10 | GE-Marquette ECG machine[9], MUSE system[10], Keras[11], TensorFlow [75], Python[12], R[13] |
| 2019 | Mohamed *et al.*[65] | 5 time-domain | 3600 | - |

Table 9 has information about the different methods used in each one of the selected studies. It is divided into the methods used for pre-processing the input data for the prediction phase and the performance evaluation. The table also includes the number of iterations used on the training as well as the data split between training and testing subsets.

[7] Microsoft Portugal, "Microsoft Excel," 2019. [Online]. Available: https://products.office.com/pt-pt/excel?legRedir=true&CorrelationId=3e4e9d3a-7d82-42a5-977c-fa3f430fa6ce&rtc=1. [Accessed: 29-Jan-2020]
[8] RapidMiner, "Lightning Fast Data Science Platform for Teams | RapidMiner©," RapidMiner, 2019. [Online]. Available: https://rapidminer.com/. [Accessed: 29-Jan-2020]
[9] "MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020]
[10] "MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020]
[11] "Home - Keras Documentation." [Online]. Available: https://keras.io/. [Accessed: 29-Jan-2020]
[12] Python Software Foundation, "Welcome to Python.org," 2001, 2019. [Online]. Available: https://www.python.org/. [Accessed: 29-Jan-2020]
[13] The R Foundation, "R: The R Project for Statistical Computing," 2018. [Online]. Available: https://www.r-project.org/. [Accessed: 29-Jan-2020]

*Table 9 - Methods applied by the selected studies.*

| Year | Study | Pre-processing method(s) | Prediction method(s) | Evaluation method(s) | Number of iterations | Data split (training/ testing) % |
|---|---|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [52] | Noise removal,<br><br>QRS detection | SVM | - | - | 47.0/53.0 |
| 2013 | Costin *et al.* [64] | Noise removal | HRV analysis and Morphologic Variability of QRS complexes | - | - | 50.0/50.0 |
| 2016 | Kim *et al.*[66] | - | CNN with ON/OFF ReLU | - | 30000 | 90.0/10.0 |
| 2016 | Shen *et al.*[62] | - | SVM | 5-fold Cross-Validation | - | - |
| 2016 | Boon *et al.*[53] | Hamilton and Tompkins algorithm, McNames algorithm | SVM | 10-fold Cross-Validation | 10 | 90.6/9.4 |
| 2017 | ElMoaqet *et al.*[50] | Noise removal | Weighted SVM | 10-fold Cross-Validation | 100 | 75.0/25.0 |
| 2018 | Rajalakshmi *et al.*[68] | Normalisation, Missing values removal | Novel Arrhythmia Fuzzy Hybrid Classifier Algorithm | - | - | - |
| 2018 | Li *et al.*[69] | Noise removal, QRS detection | Markov Chain | - | - | - |
| 2018 | Boon *et* | McNames | SVM | 10-fold | 5 | 90.6/9.4 |

| Year | Study | Pre-processing method(s) | Prediction method(s) | Evaluation method(s) | Number of iterations | Data split (training/ testing) % |
|---|---|---|---|---|---|---|
| | *al.*[63] | algorithm | | Cross-Validation | | |
| 2018 | Ebrahimzadeh *et al.* [51] | Noise removal, QRS detection | Mixture of Experts | 10-fold Cross-Validation | - | 47.0/53.0 |
| 2019 | Attia *et al.*[67] | - | CNN | - | - | 70.0/20.0 |
| 2019 | Mohamed *et al.*[65] | - | Belief Functions Theory | - | 30 | 67.0/33.0 |

The identified models/algorithms in all selected studies were compared with the reported accuracy. Some of them did not report the sensitivity and specificity, neither the F-Score nor the Area Under the Curve. Table 10 contains information about the achievements of each study.

*Table 10 - Evaluation of the selected studies.*

| Year | Study | Accuracy | Sensitivity | Specificity | F-Score | Area Under Curve |
|---|---|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [52] | 96.64% | 96.30% | 93.10% | - | - |
| 2013 | Costin *et al.* [64] | 90.00% | 89.44% | 89.29% | - | 89.40% |
| 2016 | Kim *et al.* [66] | 83.58% | - | - | - | - |
| 2016 | Shen *et al.* [62] | 75.60% | - | - | - | 83.00% |
| 2016 | Boon *et al.* [53] | 80.20% | 81.10% | 79.30% | - | - |
| 2017 | ElMoaqet *et al.* [50] | 84.90% | 66.70% | 97.00% | - | 93.50% |
| 2018 | Rajalakshmi *et al.* [68] | 82.80% | 0.40% | 0.43% | 1.21% | - |
| 2018 | Li *et al.* [69] | 82.00% | 86.00% | 80.00% | 74.51% | 90.88% |
| 2018 | Boon *et al.* | 87.70% | 86.80% | 88.70% | - | - |

| Year | Study | Accuracy | Sensitivity | Specificity | F-Score | Area Under Curve |
|---|---|---|---|---|---|---|
| | [63] | | | | | |
| 2018 | Ebrahimzadeh *et al.* [51] | 98.21% | 100.00% | 96.55% | - | - |
| 2019 | Attia *et al.* [67] | 83.30% | 82.30% | 83.40% | 45.40% | 90.00% |
| 2019 | Mohamed *et al.* [65] | 70.49% | 77.07% | 63.90% | - | - |

As Table 8 indicated, almost all of the selected studies performed feature extraction, with an exception for those who did implement a deep learning method only or did not report this information in the article. Going deeper into the analysis and comparison of the selected studies, Table 11 presents all the features selected and extracted by each one of them, ordering them referring to the number of studies using the same feature, decreasingly.

*Table 11 - Features extracted from input on each one of the selected articles.*

| Domain | Features | Studies |
|---|---|---|
| Frequency | Low-frequency band power (LF) | [52], [51], [63] |
| | High-frequency band power (HF) | [52], [51] |
| | LF/HF ratio | [64], [51] |
| | Low-frequency component of Fast Fourier Transforms (FFT-LF) | [53] |
| | High-frequency component of Fast Fourier Transforms (FFT-HF) | [53] |
| | LL-H1 | [53], [63] |
| | LL-H2 | [53] |
| | HH-H3 | [53] |
| | ROI-H1 | [53] |
| | ROI-H2 | [53] |
| | ROI-H3 | [53] |
| | QRS segment duration | [68] |
| | P-R waves interval | [68] |
| | Q-T waves interval | [68] |
| | T wave interval | [68] |
| | P wave interval | [68] |
| | Weighted centre of the bispectrum (ROI-WCOB) | [63] |

| Domain | Features | Studies |
|---|---|---|
| | Very Low-Frequency band power (VLF) | [51] |
| Time | Standard Deviation of Average of all NN interval for all 5-minute periods of the entire recording (SDANN) | [64] |
| | ST level | [62] |
| | Standard Deviation of RR intervals (SDRR) | [53], [50], [65], [51] |
| | Mean of RR intervals | [50], [65], [51] |
| | Skewness of RR intervals | [50], [65] |
| | Kurtosis of RR intervals | [50], [65] |
| | Number of adjacent RR intervals differing by more than 50 milliseconds (NN50) | [63] |
| | Sum of NN50 divided by the total number of all RR intervals (PNN50) | [63], [51] |
| | Square root of the mean of the squares of differences between adjacent RR intervals (RMSSD) | [51] |
| | Standard deviation of differences between adjacent RR intervals (SDSD) | [51] |
| | Smoothed Pseudo Winger Ville distribution (SPWVD) | [51] |
| Space | Amplitude of P wave | [62] |
| | Amplitude of Q wave | [62] |
| | Amplitude of R wave | [62] |
| | Amplitude of S wave | [62] |
| | Amplitude of T wave | [62] |
| Nonlinear | Standard Deviation 1 (SD1) | [52], [51] |
| | Standard Deviation 2 (SD2) | [52], [63], [51] |
| | SD1/SD2 ratio | [52], [51] |
| | Sample Entropy | [52], [63] |
| | Approximate Entropy | [50] |

Table 12 shows the horizon of the prediction made by every one of the selected studies, that is, in how much time can the resultant models predict AF episodes.

*Table 12 - Prediction horizon on each one of the selected articles.*

| Year | Study | Prediction horizon |
|---|---|---|
| 2012 | Mohebbi *et al.* [52] | - |
| 2013 | Costin *et al.* [64] | 30 minutes |
| 2016 | Kim *et al.*[66] | - |

| | | |
|---|---|---|
| 2016 | Shen *et al.*[62] | - |
| 2016 | Boon *et al.*[53] | 30 minutes |
| 2017 | ElMoaqet *et al.*[50] | 14 days |
| 2018 | Rajalakshmi *et al.*[68] | - |
| 2018 | Li *et al.*[69] | 2 minutes |
| 2018 | Boon *et al.*[63] | - |
| 2018 | Ebrahimzadeh *et al.* [51] | 5 minutes |
| 2019 | Attia *et al.*[67] | - |
| 2019 | Mohamed *et al.*[65] | 60 minutes |

## 3.4. Discussion

This systematic literature review aims to identify, assess and analyse the recent state-of-the-art of ECG-based models for AF Prediction using Artificial Intelligence techniques. The following paragraphs discuss the previously defined research questions.

### 3.4.1. How is the prediction problem addressed? (RQ1)

From the selected articles, most of them only address the problem of predicting AF, that is, their main focus is to predict AF and no other types of arrhythmia or heart pathologies.

All the selected articles performed classification prediction, that is, all did classify the prediction with discrete labels.

From all the twelve selected studies, only one performed a risk-based approach on the prediction problem, that is, all the other eleven did a time series prediction of AF. Regarding the number of classes used for the prediction process, only two articles reported a study using a multi-class approach, all the remaining used binary (between "PAF" and "non-PAF" events).

Despite not, all the studies reported the event horizon used for the prediction used method, two of them used a 30 minutes horizon, and the remaining used 14 days, 60 minutes, 5 minutes, 2 minutes and under a 0-minute horizon, that is, immediately before the AF event, each.

Eight out of twelve of the selected studies performed prediction of AF with input signals shorter or equal to 300 seconds, that is, five minutes long, which was also the most used length of signal in all the studies.

When looking at the datasets used by the selected articles, it is possible to see that the three most accurate models are from three of the five studies that used the dataset [60], thus identifying this as a good option for further work on assessing the problem.

### 3.4.2. What databases and features are used? (RQ2)

Despite some of the selected studies do not perform ECG signal features selection either extraction, when performing it, the selected features directly impact the model's capability of predicting AF existence with higher accuracy.

Table 11 indicates the different features selected by the articles considered in this systematic literature review.

The most used features are Standard Deviation of RR Intervals, Low-frequency band power, Mean of RR Intervals and Standard Deviation, being used by, at least, 3 different selected articles.

Most of the approaches are based solely on ECG signals, but one study combined ECG signal's data with heart morphology data. Almost half of the selected articles used the Atrial Fibrillation Prediction Database ([60]), a quarter of them used a specifically collected dataset of ECG signals. Others used a UCI Repository Warehouse ([57]) dataset, the Mayo Clinic ([58]) ECG Laboratory's database, the Medical Information Mart for Intensive Care III database ([59]), and a China Kadoorie Biobank's ([61]) database.

According to the article [68] the UCI Repository Warehouse dataset used consists of 452 instances with 279 attributes, where the ECG reports are in image format.

The Mayo Clinic ECG Laboratory's database as used by [67], included "all patients aged 18 years or older with at least one digital, normal sinus rhythm, standard 10-second, 12-lead ECG acquired in the supine position" between 1993 and 2017. The signals were acquired at a sampling rate of 500 Hz using a GE-Marquette ECG Machine[14] and stored using the MUSE data management system[15]. All the records were "over-read by a physician-supervised, trained technician, with corrections made to the diagnostic labels as needed".

As used by [65], the Medical Information Mart for Intensive Care III database was collected from 2001 to 2012. It contains information from over 40 thousand patients, about Heart Rate, Arterial Blood Pressure and Respiration. This database also contains "charts at a higher frequency like ECG and continuous blood pressure from Intensive Care Units patients". For the study, only the patients who have developed AF during their recordings are considered.

The studies [51]–[53], [63], [64] used the Atrial Fibrillation Prediction Database, which "consists of excerpts of two-channel long-term ECG (Holter) recordings and is divided into a learning set and a test set of equal size. The database includes the digitized ECG signals (sampled at 128 Hz per signal, with 12-bit resolution) and a set of unaudited, automatically-

[14] "MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020]
[15] MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020]

generated QRS annotations", as in [60]. The records were collected from 48 individuals, although the selected articles always refer to 53 or 75 participants, as in Table 7.

Finally, the study [62] was based on a database from the China Kadoorie Biobank, which is a cohort study of over 520000 adults from 10 different areas from China, collected from 2004 to 2008 using questionnaires and anthropometric and physiological measurements as well as blood samples of every participant. For the study, the 12-lead ECG data of 10 seconds duration at 500 Hz were used, which were collected from 24369 participants using a Mortara ELIx50 device during 2013 and 2014, as well as the blood pressure data (systolic and diastolic).

### 3.4.3. What pre-processing algorithms are used? (RQ3)

The pre-processing methods used in all the twelve selected studies are presented in Table 9.

Although not all the articles indicate the pre-processing methods applied, due to some of them were elaborated applying prediction methods that do not need any pre-processing of the signal, the most used pre-processing technique is Noise Removal (5 studies), followed by QRS Detection (4 studies) and Correction of Signal (2 studies). Both Normalisation and Missing Value Removal methods were applied by one study each.

### 3.4.4. What predictive algorithms are used? (RQ4)

The most used prediction method/algorithm is Support Vector Machine ([50], [52], [53], [62], [63]), followed by Convolutional Neural Network ([66], [67]).

Some other selected studies applied either statistical AI methods (HRV analysis and Morphologic Variability of QRS complexes, Belief Functions Theory), or Arrhythmia Fuzzy Hybrid Classifier, Markov Chain, or, at last, Mixture of Experts.

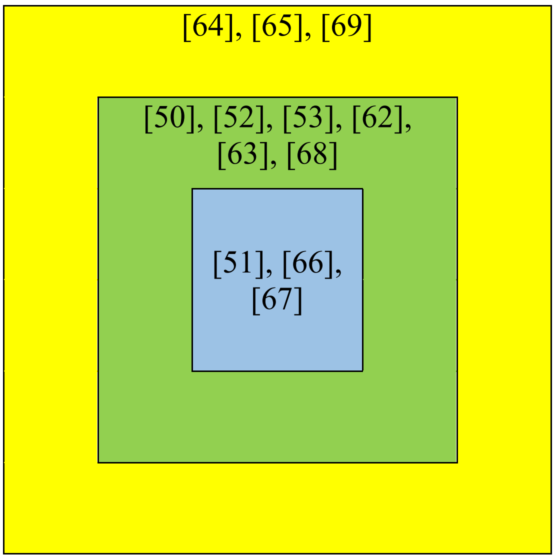



*Figure 4 - Class of Artificial Intelligence methods applied by the selected studies. Yellow=Artificial Intelligence, Green=Machine Learning, Blue=Deep Learning.*

Dividing the predictive algorithms into three classes, that is, Deep Learning, Machine Learning, and Artificial Intelligence, it is possible to identify the type of prediction approach executed by each one of the selected studies, as presented in Fig. 4.

### 3.4.5. Which are the best models that perform the best? (RQ5)

To address a comparative evaluation of the models used by the selected studies, the authors of this systematic literature review cluster the discussion in terms of:

1. Studies using the same datasets;
2. Studies applying the same prediction method or algorithm;
3. Studies based on the same input signal duration;
4. Studies within the same class of Artificial Intelligence applied method (according to Fig. 4);
5. All the studies.

1. From all the selected studies, only five of them used the same dataset, leaving all the remaining ones working with a dataset that only they used.

Thus, and comparing all the studies that used the Atrial Fibrillation Prediction Database ([60]), three of them achieved accuracies above or equal to 90.00% by applying (ordered by accuracy level decreasing) Mixture of Experts, Support Vector Machine, and Statistical AI methods ([51], [52], [64]). The two worst performing studies both used Support Vector Machine ([53], [63]), thus not being possible to indicate what was the best method to apply.

2. Regarding studies applying Support Vector Machine, those who perform the best both used as features LF, SD2 and Sample Entropy ([52], [63]).

When looking at the studies that applied Convolutional Neural Networks as a prediction method, both acquired very similar accuracy rates ([66], [67]).

3. Relatively to the studies based on the input of signals with 300 seconds length, the authors of this systematic literature review highlight the article that used as method Mixture of Experts ([51]), also linking the two best performances with the usage of the features LF, HF, SD1, SD2, SD1/SD2 and Sample Entropy ([51], [52]).

In the studies using signals of 30 seconds ([50], [66]), both performed around 85.00% of accuracy, but the second achieved higher performance, having a higher amount of individuals from whom the data was collected, as well as applying Support Vector Machine instead of Convolutional Neural Network as the first.

From the two articles that report work done with signals with 10 seconds length ([62], [67]), the second performed better than the first, and applied Convolutional Neural Network method instead of Support Vector Machine, as well as having a higher number of individuals from whom the data was collected (approximately 5.25 times).

4. From the two studies that applied Deep Learning methods ([66], [67]), both acquired very similar accuracy rates.

Looking into the articles working with Machine Learning methods (excluding those who apply Deep Learning techniques) ([50]–[53], [62], [63], [68]), the two that outperformed all the others, achieving accuracies above 95.00%, used the Atrial Fibrillation Prediction Database, worked with signals of 300 seconds long and with frequency-domain, time-domain and non-linear features extracted from the input ECG signals.

5. The results of the studies revealed that the increase in the length of the period of ECG signal sent for prediction does not necessarily increase the accuracy of the model created. The best prediction accuracies were obtained in the studies [51] (98.21%), [52] (96.64%) and [64] (90.00%), in which there were used signal parts of 300 seconds. Contrasting, the worst accuracies achieved by the models from the selected articles were obtained on the studies [65] (70.49%), [62] (75.60%) and [53] (80.20%), with signal durations of 3600, 10 and 1800 seconds respectively. These data can indicate that signals too short (10 seconds only) or too long (1800 seconds or above) are not the best approach to the problem being assessed in this systematic literature review.

At last, from the results from the three studies that applied Artificial Intelligence methods ([64], [65], [69]), the authors highlight the achieved accuracy of the first study, which worked with Atrial Fibrillation Prediction Database, having signals with 300 seconds long instead of 120 (second study) or 3600 (third study), performing better among the three.

At last, the authors highlight the achieved accuracy of the study [64], that worked with Atrial Fibrillation Prediction Database, with ECG signals 300 seconds long instead of 120 (as used on the study [65]), or 3600 (on the study [69]), performing. All these three studies used Artificial Intelligence methods.

## 3.5. Conclusion

The present systematic literature review presents and summarizes the current data-based work on predicting Atrial Fibrillation (AF) using Electrocardiogram (ECG) data as input and Artificial Intelligence (AI) methods. Twelve studies were analysed, and the main findings are summarized as follows:

- (RQ1) Despite not existing a current high number of articles published based on studies focused on prediction of AF using AI and ECG signals, most of the existing ones assess the problem by predicting only AF cases, not spending time in the prediction of other cardiovascular issues at the same time, thus being the major number of studies a binary prediction system. The higher part of the existing studies worked with ECG signals 300 seconds long, that is, five minutes. Although some studies tried increasing the length of the period of ECG signal used as input for the prediction models, it does not necessarily increase the accuracy of the obtained final model;

- (RQ2) From all the studies selected for this systematic literature review, the most accurate models were achieved using the Atrial Fibrillation Prediction Database for training. This database was also the most used, by almost half of all the selected articles. The most used features are Standard Deviation of RR Intervals, Low-Frequency band power, Mean of RR Intervals and Standard Deviation, all collected from the ECG signal inputted;
- (RQ3) Among all the selected articles, there were applied many pre-processing techniques, being the most used the Noise Removal, followed by the QRS complex detection;
- (RQ4) The trend in predictive methods based on Machine Learning techniques is increasing. From all the selected studies, the two most used methods were Support Vector Machine and Convolutional Neural Network, being the trend the Machine Learning techniques. However, the authors of this systematic literature review noticed that the usage of deep learning techniques is yet not highly accurate when comparing to simpler Support Vector Machine methods;
- (RQ5) Generally, the models based on Machine Learning methods achieved higher accuracy rates. The higher accuracy was obtained by applying a Mixture of Experts method, followed by s Support Vector Machine implementation. The selected features that conducted to higher accuracy were LF, SD2 and Sample Entropy. Also, the usage of ECG signals 300 seconds long as input for the method's training led to a high rate of prediction accuracy. The database that conducted all the three most accurate models achieved was the Atrial Fibrillation Prediction Database.

As shown by Fig. 2, between 2009 and 2019 (the period this systematic literature review covers), more than 80.00% of the total published studies were performed from 2016 ahead, 50.00% belonging to the last two years (2018 and 2019).

The amount of work on the prediction of AF episodes is rapidly increasing and showing promising results. Although deep learning methods have already shown outstanding results on the prediction of several areas, namely healthcare, but were not yet applied to many studies, that is, focusing on the prediction of AF using ECG signals. The best results tend to be achieved using Machine Learning and Deep Learning techniques, namely Support Vector Machine and Mixture of Experts.

At last, some limitations of this systematic literature review should be mentioned.

First, this systematic literature review only concerned research in papers written in English. Second, the research for articles returned few articles, even with a cross-reference of the selected studies. Third, this review excluded all the studies that included data collected from patients with recent surgical proceedings or with known cardiovascular conditions that could infer the results of an ECG exam. Finally, the selected studies had to contain evaluation measurements such as accuracy, sensitivity, specificity, or the confusion matrix, excluding any article without any of these evaluations.

# Chapter 4

# 4. Methodologies

## 4.1. Introduction

This chapter presents the methodologies and procedures performed during the present study, while theoretically introducing each one of the major three resources used, that is, the Electrocardiogram exam, the Power Spectrogram, and the Convolutional Neural Network AI model.

The main contributions of this chapter are:

1) introduce each one of the major three used resources (4.2 to 4.4);

2) present and describe the data used and the methods applied to it;

2) present the procedures of the study, while explaining and justifying each one of the several different approaches this study involved.

Part of the contents of this chapter, as well as chapter five and chapter six, were already submitted to the Computer Methods and Programs in Biomedicine scientific journal.

## 4.2. Electrocardiogram

An Electrocardiogram (ECG) signal describes the electrical activity of the heart, which relates to the impulses travelling through the heart muscles, providing information about the morphology, rhythm and heart rate. It is recorded by attaching a set of electrodes on the body surface of a patient, such as chest, legs, neck and arms [76], and its analysis has been actively used as a tool for clinical diagnosis and research.

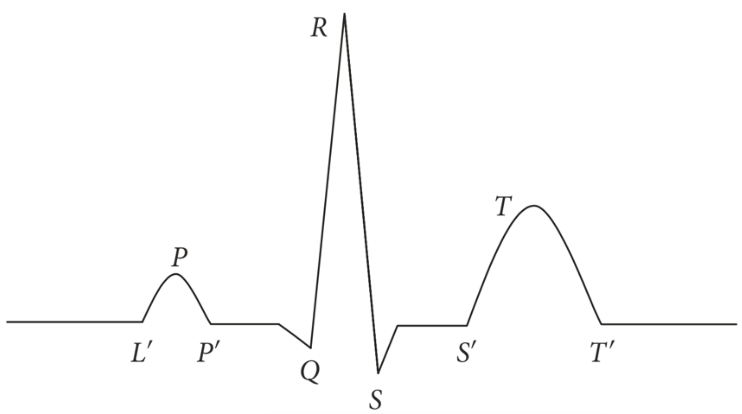



*Figure 5 - Basic shape of a healthy ECG heartbeat signal.*

The basic shape of a healthy ECG heartbeat signal is exemplified on Fig. 5. P wave reflects the sequential depolarization of the right and left atria, which is characterized by normally having positive polarity, and a duration less than 120 milliseconds, with a low frequency below 10-15 Hz. The QRS complex, that is, the region of the signal between the Q and S regions, relates to the depolarization of the right and left ventricles, lasting about 70-110 milliseconds in a normal heartbeat, and having the largest amplitude of the ECG waveforms, which is mostly concentrated between 10-40 Hz. T wave reflects ventricular repolarization and extends about 300 milliseconds after the QRS complex, with its position strongly depending on the heart rate, that is, becoming narrower and closer to the QRS complex at rapid rates [76].

Because of its ability to characterize electrophysiological as intermediate phenotypes, the ECG may be valuable for the prevention efforts along the pathway of Atrial Fibrillation (AF) [77].

Specifically on the AF prediction, the PR Interval, defined as the period between the onset of the P wave and the onset of QRS complex, with prolongation above 200 milliseconds, was shown to be associated with increased risk for AF and pacemaker implantation by [78]. From the P wave indices, that are quantitative measures of atrial electrical function derived from the ECG [79], the P wave duration quantifies the time required for atrial depolarization, which if prolonged indicates delayed intra- or interatrial conduction. This P wave prolongation has also been demonstrated to predict AF after cardiac surgery [80], and after cardioversion [81]. However, there appears to be a nonlinear relationship between this P wave index patient outcomes, and according to [82] there is a relationship between shorter P wave durations and AF. From an ECG signal, it is also possible to measure the P Wave Terminal Force (PTF) which is a reflection of left atrial activation, that is calculated by multiplying the P' duration by its amplitude [83]. This PTF has been associated with stroke [84], as well as validated as a predictor of incident AF [85]. Another marker for abnormal atrial structure was described as P Wave Area, which hence risk for AF and can be used on the AF prediction, although it is likely only being capable of predicting incident AF in the general population [85]. P Wave Axis (PWA), a reflection of anatomical features such as the positioning of the atria within the thoracic cavity and the relative size of the atria, can also reflect abnormal atrial electrical wave front propagation in a diseased myocardium, thus being a significant predictor of incident AF in United States veterans [86].

The study [87] showed that patients with a prevalence of electrocardiographic PR segment depression, when compared to controls, are at increased risk for AF as well as sudden cardiac death, meaning a short QT syndrome diagnosis. The QT interval is mainly a reflection of the time required for ventricular repolarization, which can, if prolonged, be a risk factor for sudden cardiac death and all-cause mortality [88], as well as an increased prevalence of AF when present a hereditary short QT and long QT syndromes [89], [90].

Finally, Electrocardiographic Criteria for Left Ventricular Hypertrophy (ECG-LVH) are capable of identifying individuals with higher risk index for multiple adverse outcomes, including a significant predictor of incident AF [77], [91]. Also, Premature Atrial Complexes (PACs), that is, the electrocardiographic manifestation of early atrial depolarization initiated from an outside of the sinoatrial node, have been associated with a benign ECG finding, but, when occurring frequently, they can be associated as the starting point for episodes of AF in vulnerable individuals, its presence on a 2-minute rhythm strip being a significant predictor of incident AF and stroke [92], where targeted ablation of atrial ectopy (extra beats that occur for a short time between normal regular heartbeats [93]) can reduce AF recurrence [94].

## 4.3. Power Spectrogram

A spectrogram can be defined as an intensity plot of the Short-Time Fourier Transform (STFT) magnitude, which is a sequence of Fast Fourier Transform (FFT) of windowed data segments [95]. FFT refers to an efficient implementation of the Discrete Fourier Transform (DFT), responsible for the transformation of a sequence of $N$ complex numbers into another sequence of complex number, defined by:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn} \qquad k = 0, \dots, N-1, \qquad (1)$$

where $e^{i2\pi/N}$ is a primitive $N^{th}$ root of 1.

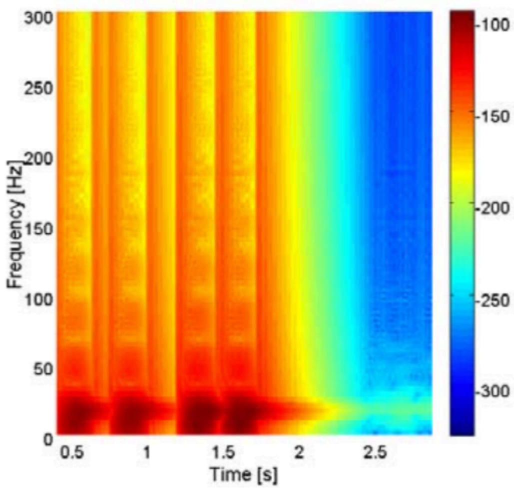



*Figure 6 - Example of a 2D representation of a spectrogram* [96].

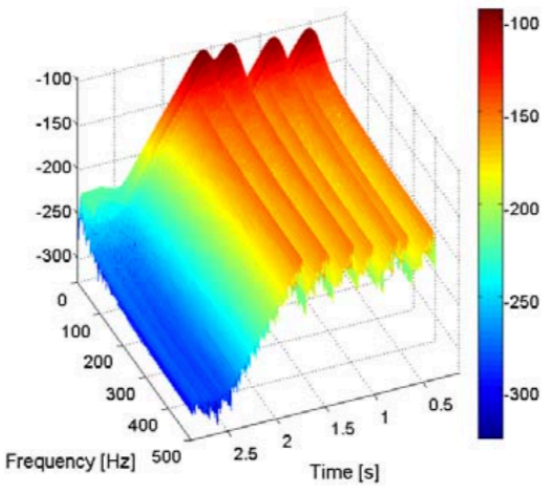



*Figure 7 - Example of a 3D representation of a spectrogram* [96].

If analysing the spectrogram as a 3D dimension image, where there is time in the X-axis, frequency of the signal in the Y-axis, then the colours of each point in the graph (Z-axis) will indicate us the amount of power the signal has in that specific frequency at a specific time, measured in decibels (dB). As presented in [96], Fig. 6 is an example of a 2D spectrogram, with a 3D representation of the same data in Fig. 7, in which it is possible to conclude that the 2D version is a flatten version of the 3D one.

## 4.4. Convolutional Neural Network

Convolutional Neural Network (CNN) is a class of deep learning models for computation of high-level representation of data by applying multiple layers of nonlinear transformations. CNN's recorded higher performance scores on object recognition [97] and classification tasks in natural images [98], having also been used in biological applications [99]–[101].

CNNs are based on a mimic of the visual information processing of the human brain by applying local filters to the input, which can be trained on the extraction of multiple local features, by applying a sliding filter across the entire visual field. This feature extraction method is known as parameter sharing and leads to a dramatic reduction in the number of parameters to train. These networks usually have in its constitution a combination of convolutional layers and local neighbourhood pooling layers, resulting in complex hierarchical representations of the inputs, which make this class of deep learning models a powerful tool in image-related applications [102].

Typically, a CNN consists of stacking multiple types of layers including convolutional, pooling and soft-max, using a set of learnable filters to provide an output classification. Each one of the applied filters have a small receptive field, and is applied in a sliding way on the input thus generating an activation map called feature map, which gives the output of those filters at every spatial position. Specifically, if the $k^{th}$ feature map at a given layer is $h^k$ and the corresponding filter contains the weight $W^k$ and the bias term $b_k$, then the output feature map $h^k$ is obtained with:

$$h_{ij}^{k} = f\big((W^{k} \cdot x)_{ij} + b_{k}\big) \qquad 2$$

where $x$ is the input vector from the previous layer and $f$ is a non-linear activation function.

Another type of layers used in CNNs is the pooling layer, which generates output feature maps by applying filters without trainable parameters, this way reducing the size of the feature map on the output and decreasing the number of parameters and computation resources in the network. This type of layer helps to achieve invariance to visual distortions by ignoring positional information.

As the final typical layer, that is, after the convolutional and pooling layers, there are two commonly used types, the soft-max and the sigmoid layers applied to image-related applications. The function of both is to transform the outputs of the previous layer to be in the range [0,1], this way providing a probabilistic meaning. The major difference between these two approaches resides on the type of image input and the type of classification that is intended.

Besides the layers that were mentioned earlier, [103] provides a full list and discussion about other types and approaches.

On the training procedure, CNNs generally are composed of two steps which are feed-forward computation and backpropagation. The first is responsible for moving forward from the input nodes, to the hidden (medium) nodes, ending at the output node(s), while the second computes a loss value between outputs of CNNd ground truth (the true classification of the input), after which it is calculated the gradient of loss functions with respect to all the weights and biases of the network, these last being updated through the obtained gradient, thus way minimizing the loss function. The widely used loss function (cross-entropy) is computed with the following equation:

$$H(p,q) = -\sum_{i=1}^{K} p_i \log(q_i) \tag{3}$$

where $p$ = [$p_1$, $p_2$, ..., $p_K$] stands for ground truth vector and $q$ = [$q_1$, $q_2$, ..., $q_K$] is for the output vector of a CNN with feed-forward computation.

Regarding the backpropagation in a CNN, there are two different ways, batch learning and online learning. Batch learning updates the weights of the CNN only after all the training samples are visited and their gradient contributions are accumulated, while online learning performs the gradient calculation after using a subset of training samples to update weights for each iteration [102]. The convergence of the learning process, that is, the time needed to reach a global minimum, is slower on the batch learning, however, the computational load is much higher applying batch learning compared to online learning. This is the main concern that allows the online learning to be very suitable and compatible with modern architectures with multiple Graphics Processing Units (GPUs).

## 4.5. Dataset

This study used the data available at the Atrial Fibrillation Prediction Database (AFPD) [60], publicly available at the Physionet Repository [104]. AFPD consists of ECG signals, each 30 minutes long, from mainly three types: "normal" tests from patients with no diagnosed PAF (file names starting with “n”); “PAF-distant” tests from patients that were PAF diagnosed but did not have a PAF onset during it (at least 45 minutes before and/or after a PAF-detected episode, when that happened) (file names starting with “p” followed by an odd number); “pre-PAF” tests from PAF-diagnosed patients where the ECG signal ends precisely before the onset of a PAF episode (file names starting with “p” followed by an even number).

All the data included in this dataset came from ECG tests of 48 different subjects, and each record contains two-channel traces sampled at 128 Hz with a 12-bit sample resolution.

From this database there were extracted 50 records of the "normal" type, plus 25 of "PAF-distant", and 25 more of "pre-PAF". In total, there were extracted 100 ECG segments, 30 minutes long each. The "normal" samples were used to evaluate our algorithm and results. These extracted records were stored in ".dat" files, which had all the required ECG points.

Each one of the 100 ECG segments has in its data both collected channels, that is, 100 samples contain data for 200 ECG segments with 30 minutes. Thus, the extracted data was demultiplexed, resulting in a total of 200 samples, 100 records of “normal” type, 50 of “PAF-distant” and 50 of “pre-PAF” types.

### 4.5.1. Data Demultiplexing

The original signal files contained in the database were interpreted using the *Python3* [105] library *WFDB* [106], which make possible the extraction of a specific channel from a single data file. As the database datafiles had two channels of signals, it was extracted the “channel=[0]” and the “channel=[1]” for every file, thus obtaining both channels separately.

## 4.6. Proposed Method

The goal of the presented study is to develop an effective and easy to implement method for the prediction of the onset of PAF. Fig. 8 shows the block diagram of the proposed method. Step 0 consists of selecting the 30-minute long signal file. Step 1 pre-processes the input ECG signal with smoothing, median filtering and notch filtering. Step 2 performs data segmentation including the slicing of the 30 minutes signal into 30 seconds portions (60 portions per 30 minutes original) and converts each 30 seconds signal into a spectrogram image. Step 3 predicts the PAF onset with a binary CNN classification.

Each one of the described blocks is detailed in the following subsections.

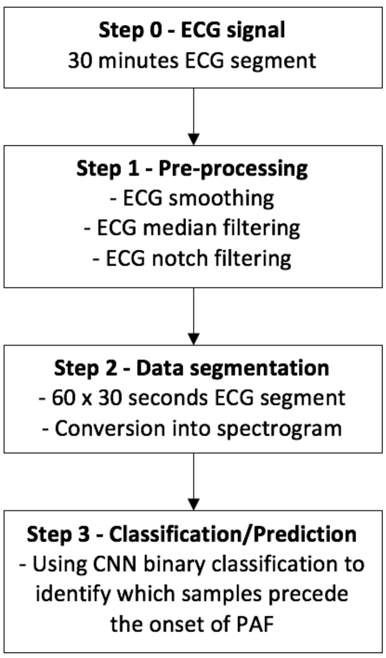



*Figure 8 - Block diagram of the proposed method.*

### 4.6.1. Pre-processing

The input signal is a 30 minutes long ECG time series. However, the collected and original data file has some noise and other interferences associated with the signal, making it necessary to process it to obtain a cleaner representation of the signal. The pre-processing stage aims to get the closest to the real data from the ECG signals' datafiles, that is, to remove the maximum possible amount of noise, interference and other phenomena that could have changed the real and true signal from the body of the patient. Fig. 9 shows the result of each one of the described steps, as applied in [51].

#### 4.6.1.1. ECG smoothing

The first step of this pre-processing stage is the removal of possible undesired values due to motion of the ECG sensors or the patient's muscle activity, with a simple moving average. It was used a window size of 3, according to the formula presented below in (1) and (2), with annotations in (3).

It was performed a comparison between a window size of 3, 5, 7 and 9, and decided to implement a size of 3 units, because it was the value that maintained the bigger amount of detail of the signal, preventing the changing of the shape of the signal.

$$sum\ of\ neighbor\ points\ (i, N) = \sum_{i-N}^{i+N} K(x) \qquad 4$$

$$smoothed\ signal\ point\ (i) = \frac{1}{2N+1}\ sum\ of\ neighbor\ points\ (i, N) \qquad 5$$

where $K$ is the list of original ECG signal points and $N$ is the parameter of the smoothing function that defines half of the size of the window of observation minus 1, that is

$$N = \frac{length\ of\ the\ observation\ window}{2} - 1. \qquad 6$$

#### 4.6.1.2. ECG median filtering

Secondly, it was applied a median filter for further reduction of the excessive noise of the original signal, with a window size of 3. The median filter is a non-linear digital filtering technique, which replaces each value with the median of the neighbouring values. Equation (4) shown the formula of this filter.

For the window option, it was performed a test between the size of 3 and 5 units, in which the lowest value did maintain the highest level of detail of the signal.

$$median\ filtered\ point\ (i, N) = median\ (i - N, \dots, i, \dots, i + N) \qquad 7$$

#### 4.6.1.3. ECG notch filtering

The third and last step of the pre-processing phase was applying a notch filter to remove the powerline interference from the ECG signal. For that, there were used several functions from the *Python3* [105] library *scipy1.4.1* [107], shown in the following equations:

$$(b, a) = iirnotch\left(\frac{cutoff}{\frac{fs}{2}}, q\right) \tag{8}$$

$$notch\ filtered\ signal = lfilter\ (b, a, original\ signal) \tag{9}$$

where *iirnotch* and *lfilter* are functions from the library *scipy*, $b$ and $a$ values are the output of the equation (8), *cutoff* is the frequency needed to remove from the signal, *fs* is the frequency of sampling of the data in use, $q$ is the order parameter of the notch filter and *original signal* is the list containing all the points of the original ECG signal to be processed and where the filter is applied.

As stated by [108], a notch filter is widely used in control systems, and, instead of a low-pass filter that attenuates all signals above a certain frequency, a notch filter only removes a narrow band of frequencies, and is also used for removing a signal resonance. In this study, it was applied this filter for removal of the frequency of 60 Hz, with the sampling frequency of the used dataset, that is, 128 Hz, and a 2nd order parameter.

Fig. 9 shows the different steps performed in this phase on an exemplary signal.

### 4.6.2. Data Segmentation

This stage was used on the preparation of the data to input to the CNN model trained in the following step, which comprised the signal cutting in shorter intervals and its conversion into spectrogram images.

#### 4.6.2.1. 60 x 30 seconds ECG segment

As presented at the start of this section, our proposed model used a CNN model training algorithm with spectrogram images as input. However, the data provided in the database were not enough for training a CNN. Thus, it was needed to increase the number of training samples while maintaining the same detail, necessary for a correct and precise prediction of the onset of PAF.

For that, the original signals were sliced, which were 30 minutes long with 128 Hz of sampling frequency, into new ones with only 30 seconds long of the same 128 Hz sampling frequency. At the end of this step, for each one of the 30 minutes signal there were 60 new 30 seconds signals, which scaled our data by 60 times, that is, 6000 samples of "normal" type, 3000 of "PAF-distant" and 3000 more of "pre-PAF". All the slicing of data process was done under a no overlap of samples condition, that is, there are no data samples repeated in these new data files of 30 seconds, providing each one of the data samples (at 128 Hz) only one time to the models.

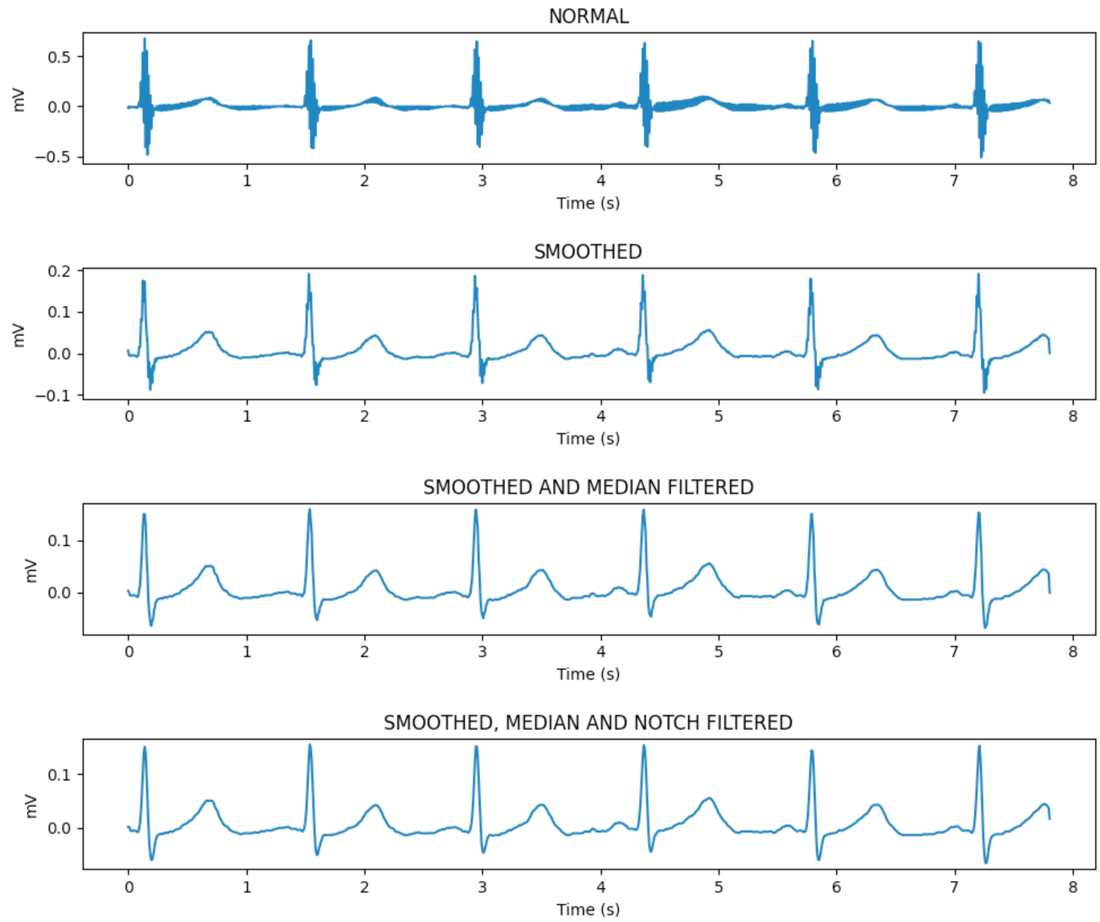



*Figure 9 - Result of each step of the pre-processing of the first seconds of the record "n01" from the database.*

The Fig. 10 shows the drawing of the first 30 seconds of the database record "p01", which were cut from the original 30 minutes long and stored separately.

### 4.6.2.2. Conversion into Spectrogram

After slicing off the database signals into 30 seconds length samples, there were generated power spectrogram images for all signal portions, without normalizing the data.

First, the spectrograms of data were generated using the *Python3* [105] function *specgram* from the library *matplotlib* [109], with the input being the ECG file data and the sampling frequency as 128 Hz. The remaining parameters were not changed, the default values for each one of them being set to 256, including the number of data points used in each block for the *Fast Fourier Transform*.

Fig. 11 shows the produced spectrogram for the wave presented in Fig. 10, that is, the first 30 seconds from the signal "p01" from the database.

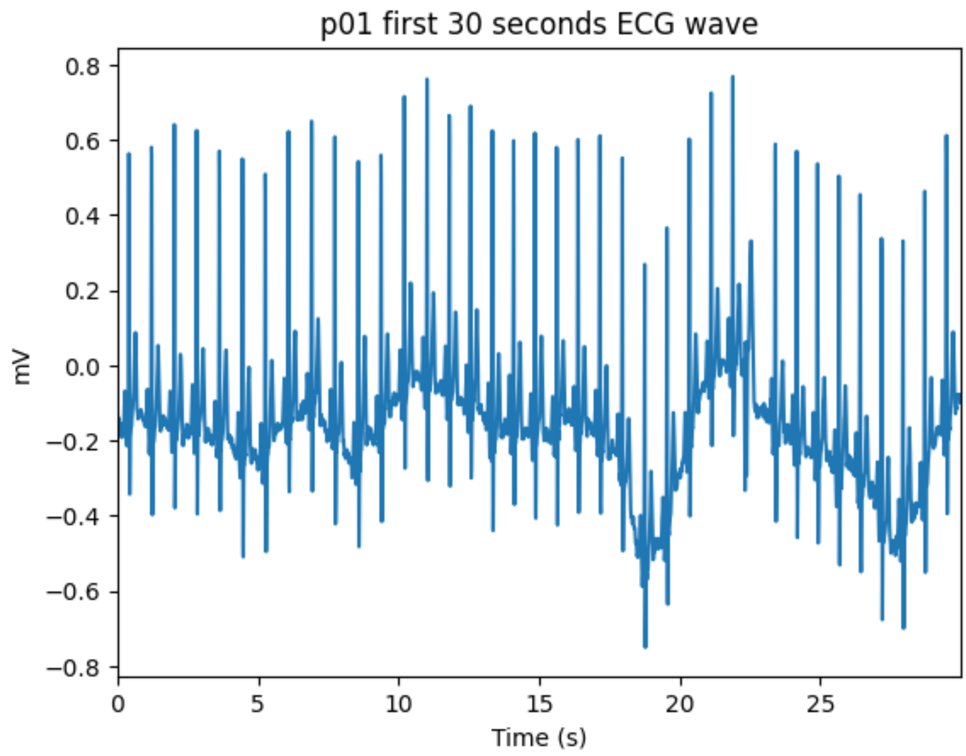



*Figure 10 - Wave signal of the first 30 seconds of the record "p01" from the database.*

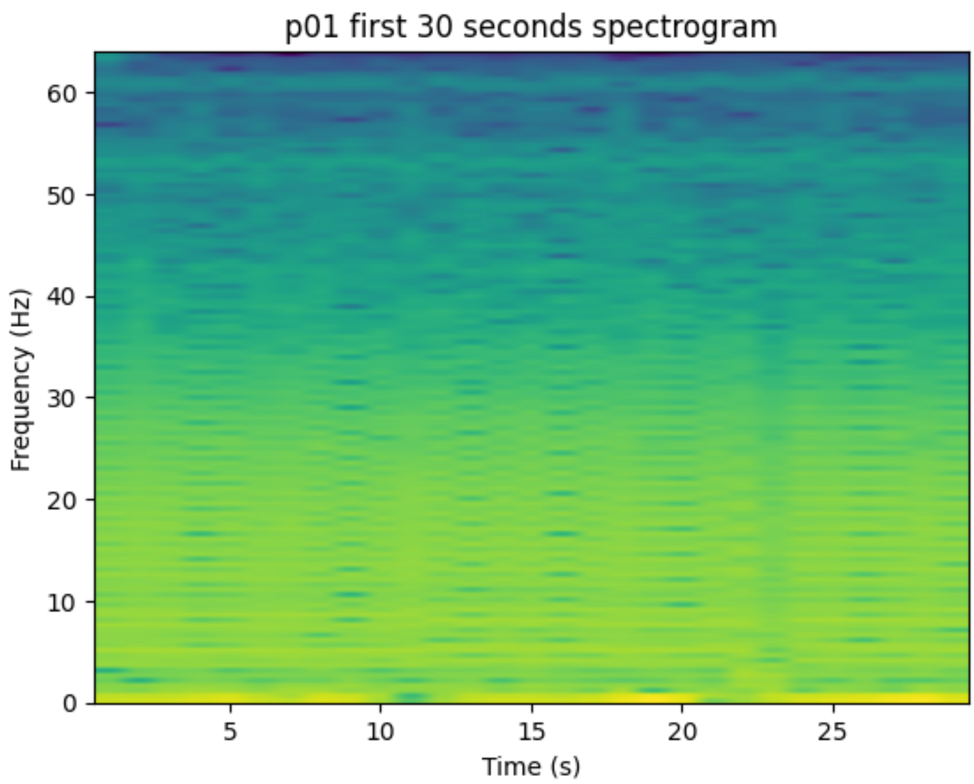



*Figure 11 - Spectrogram of the signal of the first 30 seconds of the record "p01" from the database.*

In it, it is possible to have a better understanding of the signal in need to analyse, by plotting the power of the signal depending on the frequency at each instance of time. The colour bar used in this approach represented the lowest power with a dark blue, and the highest with a bright yellow colour.

This way, it was possible to assure that the first 30 seconds of the record “p01” of the database have a predominance of signals with low frequency, that is, it is easily seen that the majority of the yellow regions on the spectrogram are below the frequency of 10 Hz in the 30 seconds signal length.

However, in Fig. 12 and Fig. 13, it is possible to compare the spectrogram images of the last 30 seconds of the records “n01” and “p01”, respectively, that is, the last seconds before a normal condition ECG and a PAF episode. It is easily noticeable the difference between the yellow regions between both images, with the majority of signal power on low frequencies for a pre-PAF episode and medium-high frequencies for a normal ECG signal. Fig. 14 and 15 also follow this trend, representing the first 30 seconds of the records “n01” and “p01”, respectively.

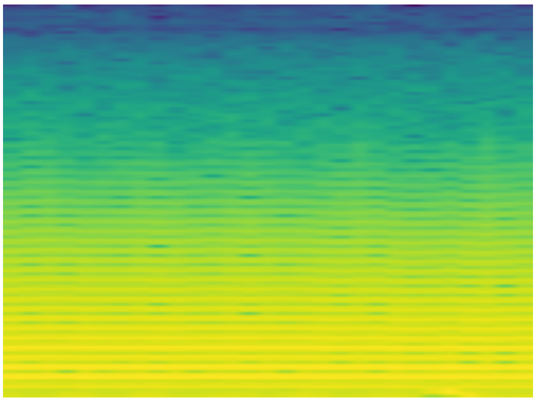


*Figure 12 - Spectrogram of the last 30 seconds of the record "n01".*

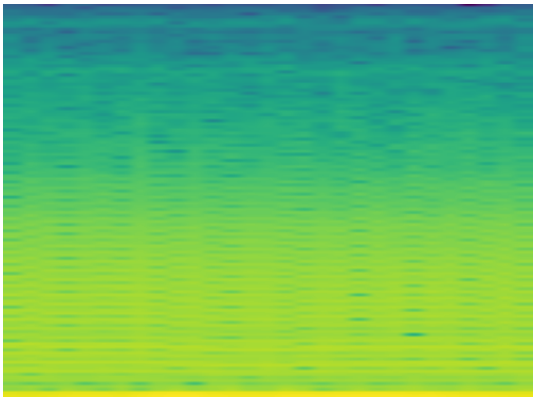


*Figure 13 - Spectrogram of the last 30 seconds of the record "p01".*

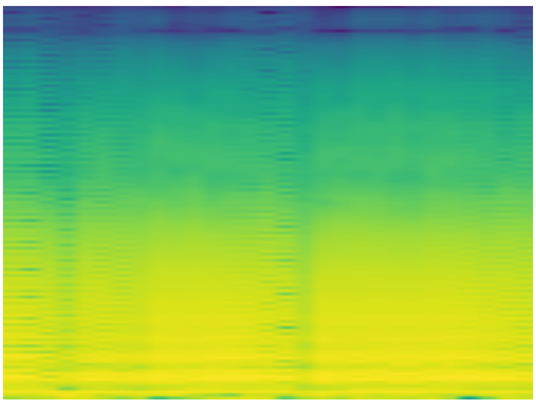


*Figure 14 - Spectrogram of the first 30 seconds of the record "n01".*

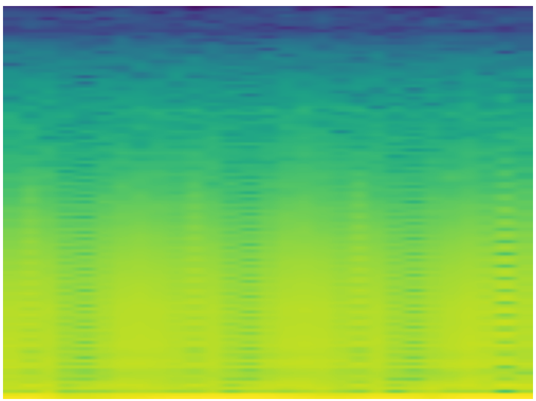


*Figure 15 - Spectrogram of the first 30 seconds of the record "p01".*

All the figures represent the data as it is at the end of step 2 of the pre-processing phase, that is, ready to be fed to the Convolutional Neural Network described at step 3. For this reason, there are no axis or white spaces on them, only the spectrogram graph.

### 4.6.3. Classification/Prediction

The next step was constructing, training and testing the CNN model for the classification of the input data. For the prediction of the onset of a PAF episode, our model was trained to classify between samples that preceded or not a PAF event, that is, learning the differences and patterns in data preceding a PAF episode ("pre-PAF" type or "PAF-distant" type) and data not preceding, in a short or long time interval, a PAF episode ("normal" type).

Despite the slicing of the data coming from the used dataset, the resulting amount of data from "pre-PAF" and "normal" types was not enough for a high accuracy as expected initially, which led us onto three different approaches, organized by two main types, "Short Prediction" (SP) and "Long Prediction" (LP), which refers to the prediction horizon of each one, 29.5 minutes and >45.0 minutes, respectively, as stated by Table 13.

*Table 13 - Organization and types of the different approaches.*

| Main type of the approach | Designation of the approach | Prediction horizon |
|---|---|---|
| Short Prediction (SP) | SP Simple Approach (SPSA) | 29.5 minutes |
| Short Prediction (SP) | SP Hybrid Approach (SPHA) | 29.5 minutes |
| Long Prediction (LP) | LP Simple Approach (LPSA) | >45.0 minutes |

The LP Simple Approach (LPSA) was trained, validated and tested with all the extracted data from the dataset, that is, it was developed using all the "pre-PAF", "PAF-distant" and "normal" data.

The SP Simple Approach (SPSA) was conducted by training and validating the model with only "pre-PAF" and "normal" data, and testing also with only these two types.

In the remaining approach, SP Hybrid Approach (SPHA), it was trained and validated a pre-trained model with the same data types as on SPSA, that is, "pre-PAF" and "normal" data, and tested with this same data types. The pre-trained model used by this approach was the best result selected from the LPSA.

All the three techniques are detailed as follows and were implemented using a CNN with the *Keras Framework* [110] and *Python3* [105].

The data contained in the testing phase was never presented to the model during the training phase, that is, the model only accessed this new data only for testing. This guaranteed the model is robust and can generalise well. In the separation of the training and testing data sets, this was performed using an automated and random selection procedure coded with *Python3* [105] programming language, thus assuring the models had access to test data from any possible 30 seconds segment of the original ECG signals.

In all approaches, all the input data had the same image size of 496 by 369, and all were trained and tested 10 times with a batch size of 130 and 100 fixed epochs. The obtained results are described in Section 5. A diagram of the network used in all the three approaches is presented by Fig. 19 for the SPSA and by Fig. 20 for both the SPHA and LPSA.

#### 4.6.3.1. Short Prediction Simple Approach (SPSA)

This approach was conducted with three layers of convolution, with 32, 32 and 64 filters by this specific order, and a kernel size of (2,2), ReLU as the activation function and with max-pooling of (2,2) at the end of each one of the layers. This step is known as the Feature Extraction phase of the CNN model, during which the algorithm selects the best features that allows distinguishing, in this specific case, between a "normal" and a "pre-PAF" ECG signal's spectrogram image.

These layers were followed by a flatten instruction, which converts the matrix resulting from the convolution layers into a single array with all the data from these.

In the Classification phase, a fully connected layer with 64 neurons was added with a ReLU function as activation, followed by a dropout of 0.5. The fully connected, or dense, layer was responsible for measuring the importance of each one of the selected features in the classification process, thus adjusting the algorithm to the training data, and the dropout prevented the model from overfitting the training samples, thus generalising better when classifying data different from the training one.

At the end of the model architecture, it was used a fully connected layer with only one neuron, and applied a Sigmoid function as activation, thus returning a probability of a spectrogram being “pre-PAF” type or not, so that values above 0.5 represent a binary true classification for ECG signals preceding a PAF episode in the next 29.0 minutes and 30 seconds or less. The proposed model diagram is represented in Fig. 16.

For compiling this training model, it was used the Binary Crossentropy function as loss function and *rmsprop* as optimiser parameter. On the train and test data sets, it was applied a rescale of 1/255, and a shear range of 0.2, a zoom range of 0.3 and a true horizontal flip for the training data.

This algorithm was fed with 6704 data samples for training (4454 of “normal” type and 2250 of “pre-PAF” type) and 2235 samples for testing (1485 “normal” and 750 “pre-PAF” type), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing.

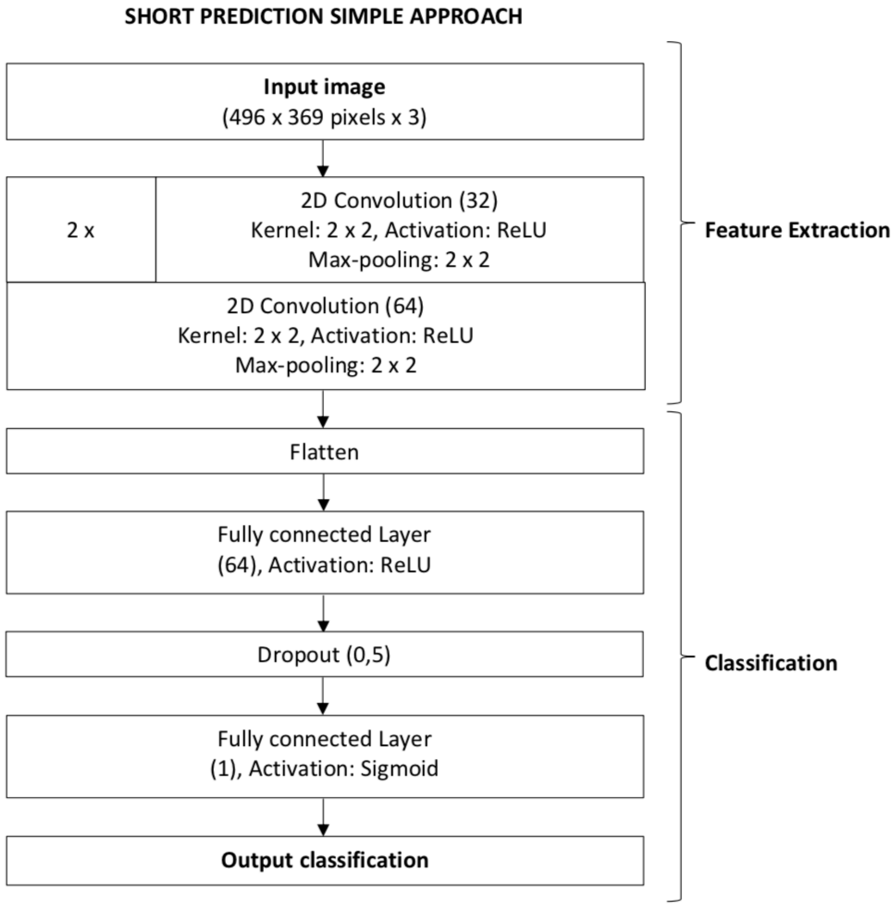



*Figure 16 - Block Diagram of the SPSA model.*

#### 4.6.3.2. Short Prediction Hybrid Approach (SPHA)

This hybrid approach was implemented with almost the same specifications as the previous method, except for the number of samples used and the model architecture.

In this approach, it was performed an attempt to improve the results achieved by training and testing a pre-trained model from the LPSA only with "normal" and "pre-PAF" data types. The pre-trained model, inserted at the start of the training process for this approach, was trained and tested with more data ("normal" type as the negative class for PAF prediction, "pre-PAF" and "PAF-distant" types as positive). Then, this approach model itself was trained and tested with only "normal" and "pre-PAF" data, thus facing the low amount of training data issue and validating the model with the same data as on the simple approach, only by using a pre-trained model from the LPSA. This allowed us to classify an ECG section spectrogram between positive and negative for a preceding PAF episode in the next 29.0 minutes and 30 seconds, or less.

On the construction of this algorithm, the SPHA, the model was designed as having three layers of convolution, with 32, 32 and 64 filters by this specific order, and a kernel size of (2,2), ReLU as the activation function and with max-pooling of (2,2) at the end of each one of the layers on the Feature Selection phase. Next, it was coded a flatten instruction (as on the SPSA) and six fully connected layers with 64 neurons with a ReLU function as activation, followed by a dropout of 0.5.

As the final step, and as on the simple approach, it was implemented a fully connected layer with only one neuron, and applied a Sigmoid function as activation, thus returning a probability of a spectrogram being “pre-PAF” type or not, assuming a true case when the probability goes above 0.5. The proposed model diagram is represented in Fig. 17.

This algorithm was fed with 6704 data samples for training (4454 of “normal” type and 2250 of “pre-PAF” type) and 2235 samples for testing (1485 “normal” and 750 “pre-PAF” type), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing.

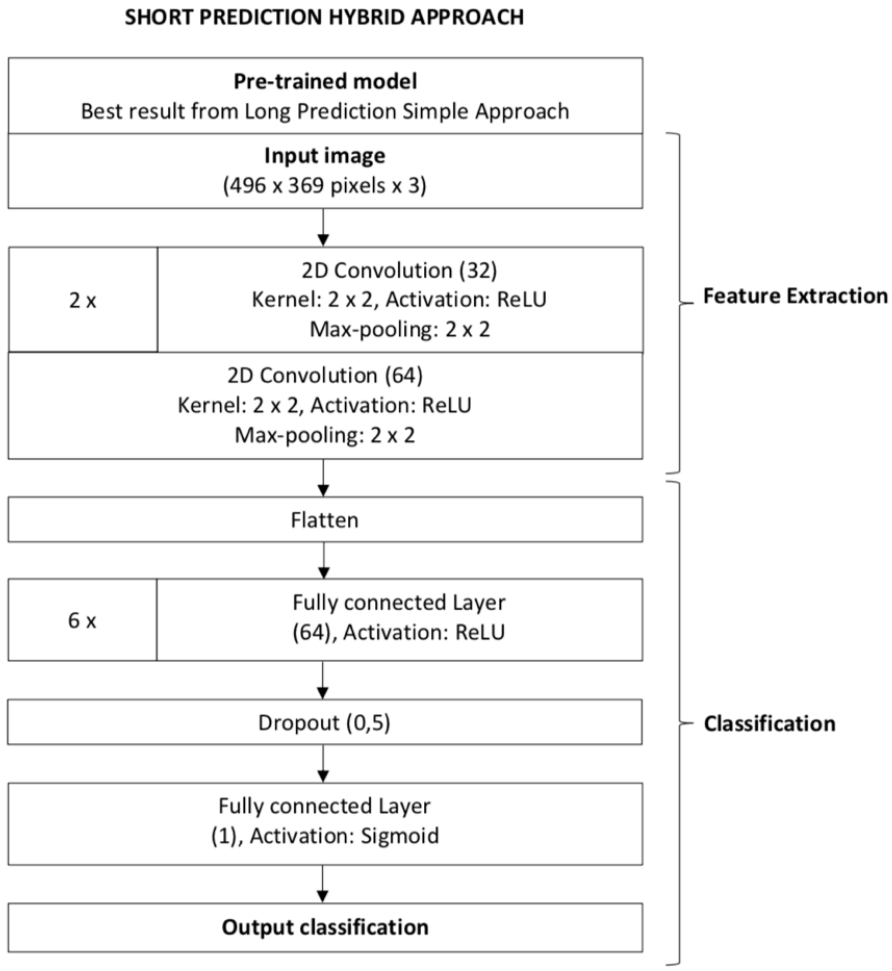



*Figure 17 - Block Diagram of the SPHA model.*

#### 4.6.3.3. Long Prediction Simple Approach (LPSA)

This final approach was implemented with the same network specifications as SPHA, except for the number of samples used and the non-usage of a pre-trained model at the start, that is, it was trained from scratch, as well as SPSA.

The algorithm was fed with 8954 data samples for training (4454 of “normal” type and 4500 of “pre-PAF” and “PAF-distant” types combined) and 2985 samples for testing (1485 “normal” and 1500 of “pre-PAF” and “PAF-distant” types combined), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing.

By training this model with all the available and extracted data from the dataset ("normal" type as the negative class for PAF prediction, "pre-PAF" and "PAF-distant" types as positive), this allowed us to classify an ECG section spectrogram between positive and negative for a preceding PAF episode with an event horizon of 45.0 minutes or more, that is, being able to predict the onset of a PAF event in the next 45.0 minutes, at least. The proposed model diagram is represented in Fig. 18.

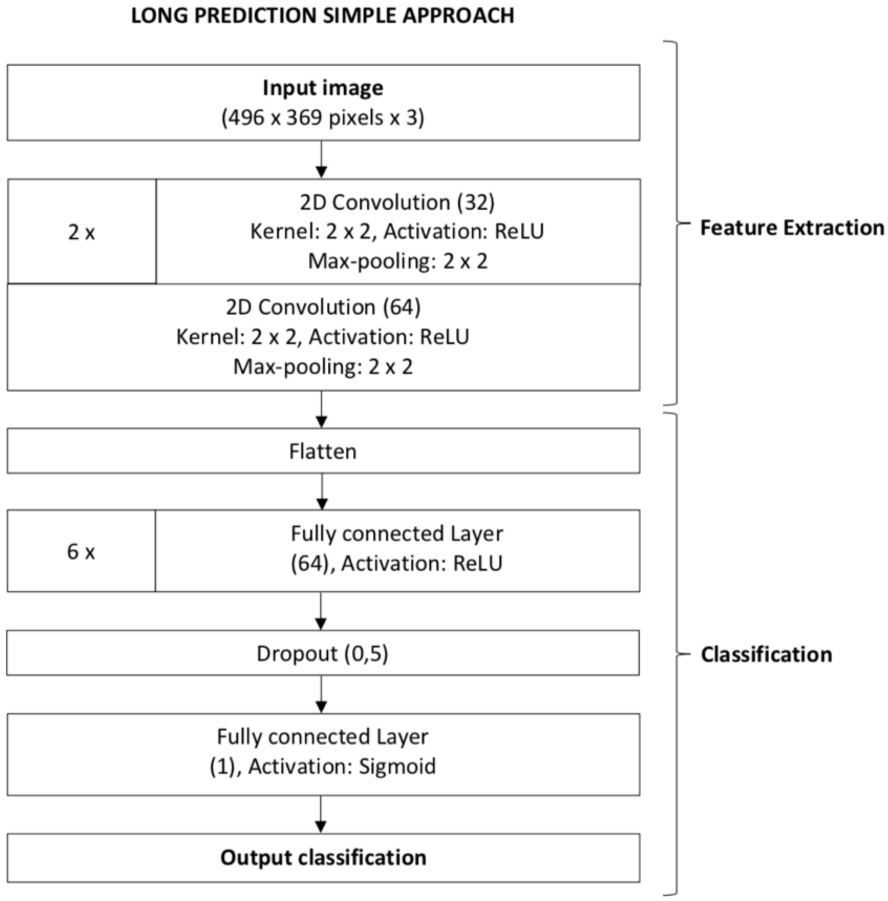



*Figure 18 - Block Diagram of the LPSA model.*

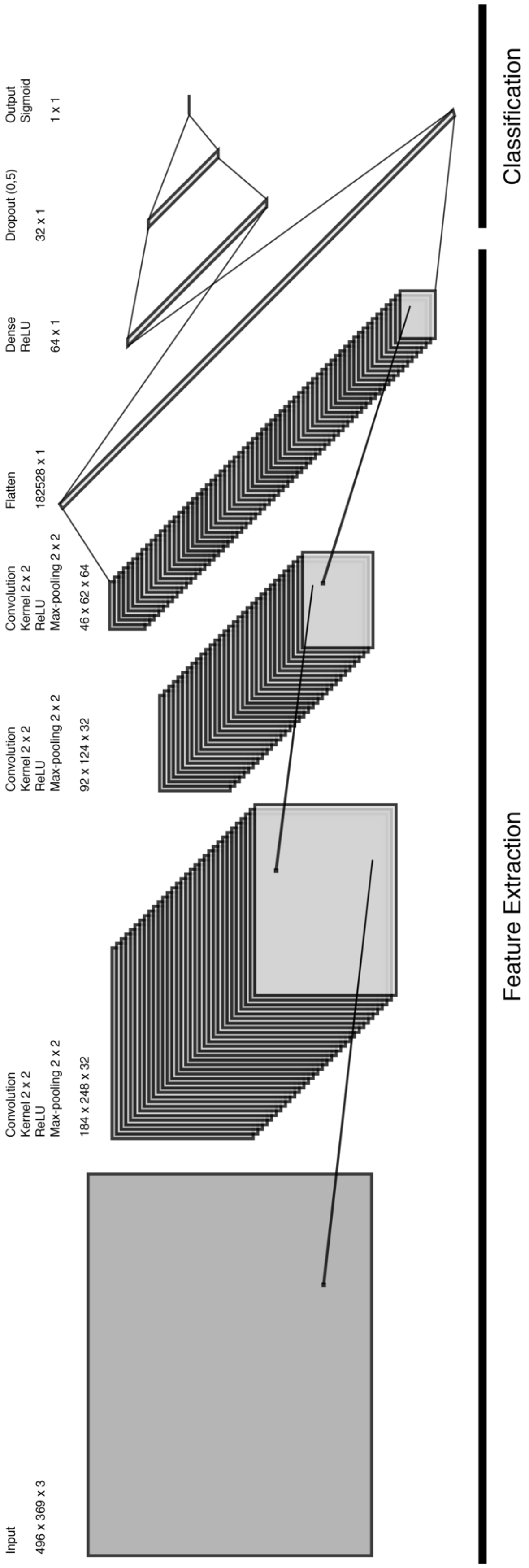



*Figure 19 - Diagram of the SPSA model.*

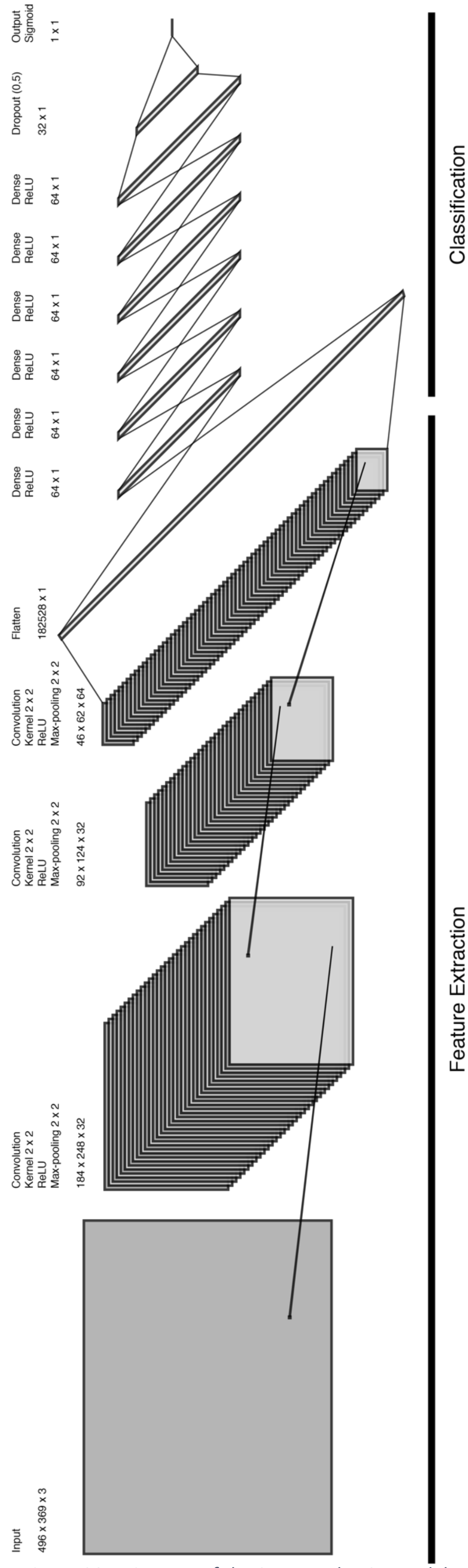



*Figure 20 - Diagram of the SPHA and LPSA models.*

# Chapter 5

# 5. Results

## 5.1. Introduction

This chapter presents the results obtained during the present study, also introducing and detailing the metrics used to evaluate them.

To evaluate the performance of the proposed methods and obtained classification models, there were used the following four measures:

$$Accuracy\ (\%) = \frac{TN + TP}{TN + FP + FN + TP} \times 100 \qquad 10$$

$$Sensitivity\ (\%) = \frac{TP}{TP + FN} \times 100 \qquad 11$$

$$Specificity\ (\%) = \frac{TN}{TN + FP} \times 100 \qquad 12$$

$$F1\ score\ (\%) = 2\left(\frac{Precision\ \times\ Recall}{Precision + Recall}\right) \times 100 \qquad 13$$

where *TN*, *TP*, *FP* and *FN* stand for True Negative, True Positive, False Positive and False Negative, respectively, and

$$Precision\ = \frac{TP}{TP + FP} \qquad 14$$

$$Recall\ = \frac{TP}{TP + FN}. \qquad 15$$

If for example a “pre-PAF” episode is classified as a “pre-PAF” episode, then it is said that the episode is classified as *TP*. On the other hand, if a "normal" episode is classified as "normal", then it is a *TN* case. Any "normal" episode classified as "pre-PAF" by mistake will produce an *FP* episode, as well as a "pre-PAF" episode classified as "normal" will produce an *FN* episode.

Also, there were used four additional and specific types of metrics, Validation Loss (val_loss), Validation Accuracy (val_acc), Loss and Accuracy (acc).

Accuracy and Validation Accuracy are both accuracies but measured according to two methods: while the first is the accuracy of the model on correctly classifying the data with which it was trained (data that belongs to the training data portion), the second measures the accuracy of the model on classifying the test data, which it has never "seen" before, that is, was not available during the training phase (only belongs to the test data portion).

During the training and test phases, the model is tested to classify each one of the data samples that are available in that step, that is, after every epoch of training the model is tested to find its capability for correctly classify all the samples. After trying to classify all the samples available (only the train samples during the training phase, and only the test samples during the test phase), it is measured the loss, which is a measurement of the difference between the classification given by the model and its true class. This way, Validation Loss refers to the loss value on each one of the training epochs, and Loss is the loss value after the testing phase.

## 5.2. Prediction of PAF onset

In Table 14 the results achieved by each one of the related work studies are presented, in comparison to the results obtained in this research. The metrics resulting of this study were evaluated ten times with the same specifications of the model train and test phases, having achieved an average of 95.98% of accuracy, 91.09% of sensitivity, 98.45% of specificity and 93.82% of F1 score for the SPHA, 85.01% of accuracy, 67.29% of sensitivity, 93.95% of specificity and 74.77% of F1 score for the SPSA, and 94.22% of accuracy, 94.40% of sensitivity, 94.05% of specificity and 94.26% of F1 score for the LPSA.

The metrics for this research in Table 14 present the average results out of 10 training and testing sequences, followed by the worst and best-measured metric in the respective type.

*Table 14 - Summary of different methods for PAF onset prediction and their reported results, compared to this study average metrics.*

| Author | Method | Accuracy | Sensitivity | Specificity | F1-Score |
|---|---|---|---|---|---|
| Mohebbi *et al.* (2012) [52] | SVM | 96.64% | 96.30% | 93.10% | - |
| Costin *et al.* (2013) [64] | HRV analysis and Morphologic Variability of QRS complexes | 90.00% | 89.44% | 89.29% | - |
| Kim *et al.* (2016) [66] | CNN with On/Off ReLU | 83.58% | - | - | - |
| Shen *et al.* (2016) [62] | SVM | 75.60% | - | - | - |
| Boon *et al.* (2016) [53] | SVM | 80.20% | 81.10% | 79.30% | - |
| ElMoaqet *et al.* (2017) [50] | Weighted SVM | 84.90% | 66.70% | 97.00% | - |
| Rajalakshmi *et al.* (2018) [68] | ARFC algorithm | 82.80% | 0.40% | 0.43% | 1.21% |
| Li *et al.* (2018) [69] | Markov Chain | 82.00% | 86.00% | 80.00% | 74.51% |
| Boon *et al.* (2018) [63] | SVM | 87.70% | 86.80% | 88.70% | - |
| Ebrahimzadeh *et* | Mixture of | 98.21% | 100.00% | 96.55% | - |

| | | | | | |
|---|---|---|---|---|---|
| *al*. (2018) [51] | Experts | | | | |
| Attia *et al*. (2019) [67] | CNN | 83.30% | 82.30% | 83.40% | 45.40% |
| Mohamed *et al*. (2019) [65] | Belief Functions Theory | 70.49% | 77.07% | 63.90% | - |
| **This research (SPSA)** | **CNN** | **85.01%** (81.61% - 88.19%) | **67.29%** (47.47% - 82.80%) | **93.95%** (85.86% - 98.86%) | **74.77%** (63.40% - 80.95%) |
| **This research (SPHA)** | **CNN (pre-trained with different data)** | **95.98%** (94.68% - 96.73%) | **91.09%** (86.40% - 93.60%) | **98.45%** (96.84% - 99.26%) | **93.82%** (91.59% - 95.06%) |
| **This research (LPSA)** | **CNN** | **94.22%** (91.66% - 96.52%) | **94.40%** (90.27% - 97.40%) | **94.05%** (88.48% - 96.84%) | **94.26%** (91.95% - 96.56%) |

Fig. 21, 22 and 23 present the Loss, Accuracy, Validation Loss and Validation Accuracy curves of the best three obtained models, that is, of the SPSA, SPHA and LPSA approaches (only the best model achieved, by comparing accuracy and loss). In them, it is possible to visualize the evolution of the growing of accuracy and decreasing of loss along the training and testing phases of each model. The curves associated to the training phase, *loss* and *acc*, are represented in green colour, with dashed and continuous lines, respectively. The curves of *val_loss* and *val_acc*, from the test phase, assume a dark yellow colour, in dashed and continuous lines, respectively.

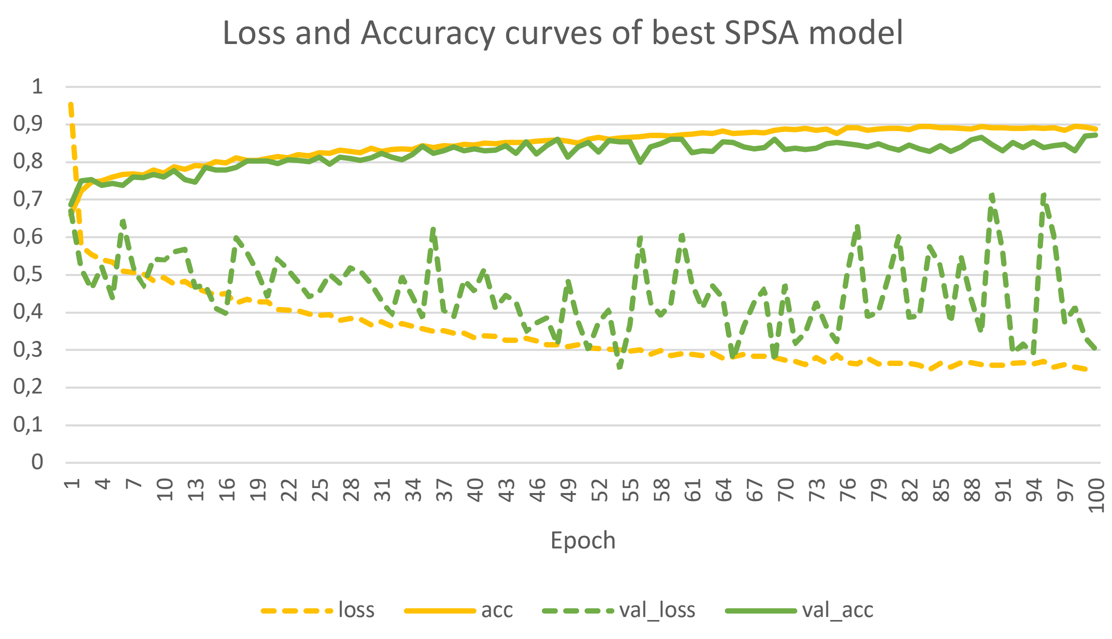



*Figure 21 - Loss and Accuracy (train and test) curves of best SPSA model.*

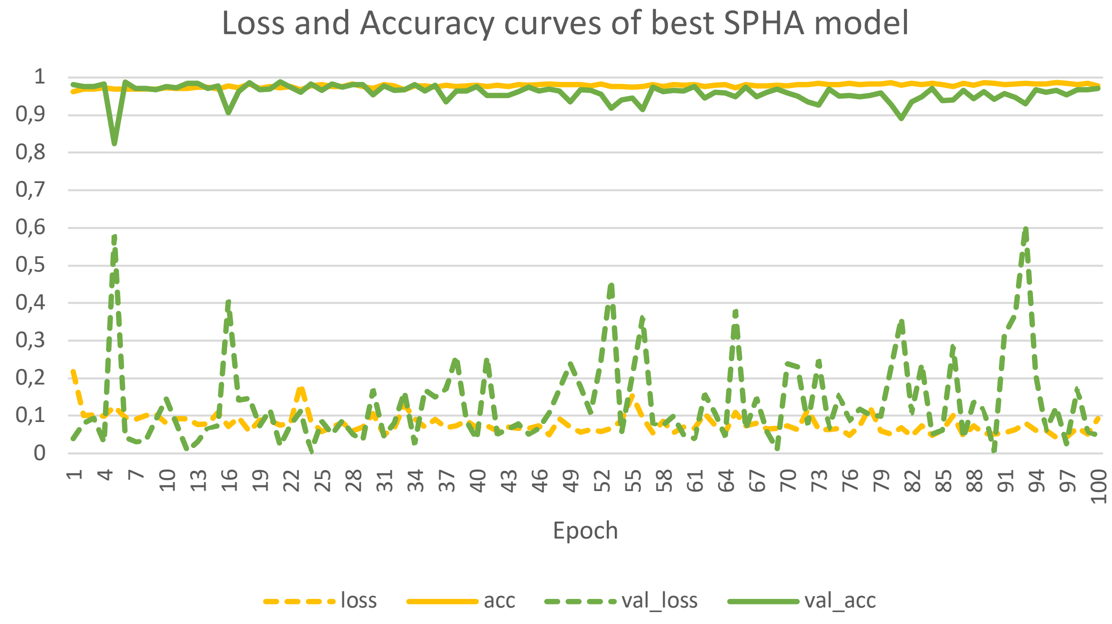



*Figure 22 - Loss and Accuracy (train and test) curves of best SPHA model.*

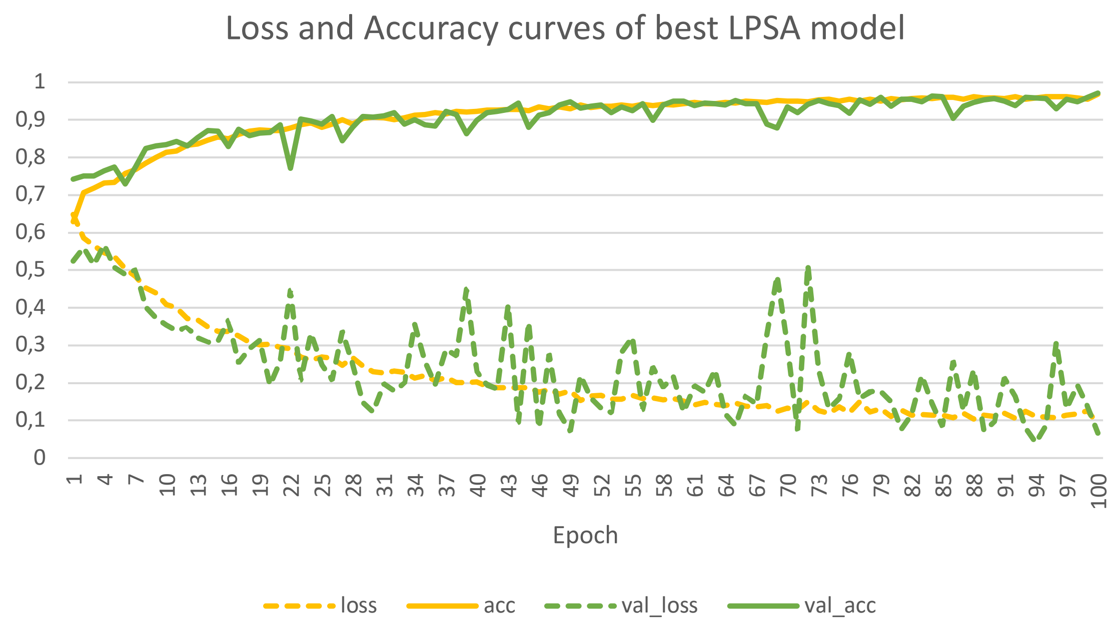



*Figure 23 - Loss and Accuracy (train and test) curves of best LPSA model.*

# Chapter 6

# 6. Conclusions

## 6.1. Introduction

This chapter is intended to discuss the results achieved during the work, as well as to conclude its content and present some new approaches that can be undertaken by future work on the topic.

It is organized in three parts, the first one for the introduction and the remaining for the two topics above mentioned.

## 6.2. Discussion

With the results of this study, it is possible to conclude that an ECG of 30 seconds is an appropriate length of a data sample to predict the onset of a PAF episode in a patient. It is also possible to conclude that the spectrogram representation of these signals can maintain the required detail for a neural network to detect some patterns and recognise the difference between a "normal" ECG and a record that will lead to a PAF episode in 29.0 minutes and 30 seconds or less, or even 45.0 minutes. This can reduce the computational power required to these technological health approaches, compared to a raw and fully detailed ECG signal, which can overfit and reduce the precision of such algorithms.

The resulting models from this work are three examples of how technology can help to diagnose and prevent several health conditions ahead of time.

The algorithms SPSA and SPHA can be applied, for example, in an Intensive Care Unit (ICU) tool to alert the physicians for possible dangerous situations happening with a patient. They can also be used to prevent and reduce False ICU alarms, similar to [111]. On the other hand, the algorithm LPSA opens the possibility of an AF health condition diagnosis in a patient without previous AF episodes, by providing a PAF episode prediction for several minutes or hours in advance, according to the data used in its development.

This study achieved very high accuracy and precision on the prediction of a PAF episode compared to the previous related works. The proposed models were only slightly worse than [51], which is not very detailed and is challenging to replicate. Also, this study results don't need the constant adjustment of parameters as it is necessary for statistical methods that are being applied nowadays.

However, this study has some limitations that should be mentioned. First, the amount of data available and selected for this particular approach was not large enough for deep learning models. However, the applied data augmentation technique of cutting the original signals into shorter segments proved to be effective. Also, the used dataset can contain errors, kindly reported by the authors in [112]. The conditions of the data labelling were not clear enough and can contain errors, such as some "normal" data samples can be a result of a labelling mistake, coming from an undiagnosed PAF patient or having some portions of PAF related features that were ignored or mistaken during the classification by the physician in charge.

## 6.3. Conclusion and Future Work

Three practical and precise deep learning-based algorithms for predicting the onset of a PAF attack were presented in this study. The achieved high accuracy, specificity and sensitivity were obtained by combining some simple, yet effective and reliable noise reduction techniques and innovative use of spectrogram images for the conversion of its resulting ECG exam signals.

One possible way to further improve this research topic would be, at first, on increasing the amount of data available for the models to train, by merely collecting new ones or by using other datasets. Likewise, from these new datasets could cut the ECG portions immediately before the onset of a PAF episode, which could address for one of the most significant limitations of this work, due to the enormous amount of data related to PAF episodes detection, that is, ECG data containing the PAF episode itself and some signal before or after it. Furthermore, the three proposed approaches open the possibility of combining other algorithms with these pre-trained models. This could lead to a broader generalisation of the resulting method, which is one of the most significant limitations on using deep learning algorithms, that is, algorithms that easily recognise and detect patterns on the data in which they were trained but find it very hard to generalise for new and unseen data.

The obtained results of this study demonstrate that its methods can be used as an accurate and reliable tool for the prediction of the onset of PAF events.

# Bibliography


[1] WHO, “WHO | About cardiovascular diseases,” *WHO*, 2011. [Online]. Available: https://www.who.int/cardiovascular_diseases/about_cvd/en/%0Ahttp://www.who.int/mediacentre/factsheets/fs317/en/%0Ahttp://www.who.int/cardiovascular_diseases/about_cvd/en/. [Accessed: 29-Jan-2020].

[2] W. H. Organization, “Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016,” Geneva, 2018.

[3] E. Wilkins, W. L., K. Wickramasinghe, and B. P, “European Cardiovascular Disease Statistics 2017 edition,” *Eur. Hear. Netw.*, pp. 8–15; 94, 118, 127, 149, 162, 174, 2017.

[4] J. Leal, R. Luengo-Fernández, A. Gray, S. Petersen, and M. Rayner, “Economic burden of cardiovascular diseases in the enlarged European Union,” *Eur. Heart J.*, vol. 27, no. 13, pp. 1610–1619, Jul. 2006, doi: 10.1093/eurheartj/ehi733.

[5] D. J. Gladstone *et al.*, “Atrial fibrillation in patients with cryptogenic stroke,” *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2467–2477, Jun. 2014, doi: 10.1056/NEJMoa1311376.

[6] T. Sanna *et al.*, “Cryptogenic stroke and underlying atrial fibrillation,” *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2478–2486, Jun. 2014, doi: 10.1056/NEJMoa1313600.

[7] Mayo Clinic, “Atrial fibrillation - Symptoms and causes - Mayo Clinic,” 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624. [Accessed: 29-Jan-2020].

[8] Mayo Clinic Staff, “Electrocardiogram (ECG or EKG) - Mayo Clinic,” *Mayo Clinic*, 2018. [Online]. Available: https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983. [Accessed: 29-Jan-2020].

[9] I. Matias, N. Pombo, and N. Garcia, “Towards a fully automated bracelet for health emergency solution,” in *IoTBDS 2018 - Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security*, 2018, vol. 2018-March, pp. 307–314, doi: 10.5220/0006800803070314.

[10] N. Pombo, P. Araújo, J. Viana, and M. D. Da Costa, “Evaluation of a ubiquitous and interoperable computerised system for remote monitoring of ambulatory post-operative pain: A randomised controlled trial,” *Technol. Heal. Care*, vol. 22, no. 1, pp. 63–75, Jan. 2014, doi: 10.3233/THC-130774.

[11] N. Pombo and N. M. Garcia, “UbiSleep: An ubiquitous sensor system for sleep monitoring,” in *International Conference on Wireless and Mobile Computing, Networking and Communications*, 2016, doi: 10.1109/WiMOB.2016.7763192.

[12] N. Pombo, P. Araújo, J. Viana, B. Junior, and R. Serrano, “Contribution of web services to improve pain diaries experience,” in *Lecture Notes in Engineering and Computer Science*, 2012, vol. 2195, pp. 589–592.

[13] H. P. Da Silva, J. Guerreiro, A. Lourenço, A. Fred, and R. Martins, “BITalino: A novel hardware framework for physiological computing,” in *PhyCS 2014 - Proceedings of the*

*International Conference on Physiological Computing Systems*, 2014, pp. 246–253, doi: 10.5220/0004727802460253.

[14] T. Lysaght, H. Y. Lim, V. Xafis, and K. Y. Ngiam, "AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research," *Asian Bioeth. Rev.*, vol. 11, no. 3, pp. 299–314, Sep. 2019, doi: 10.1007/s41649-019-00096-0.

[15] D. E. Singer, G. W. Albers, J. E. Dalen, A. S. Go, J. L. Halperin, and W. J. Manning, "Antithrombotic therapy in atrial fibrillation: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy," in *Chest*, 2004, vol. 126, no. 3 SUPPL., pp. 429S-456S, doi: 10.1378/chest.126.3_suppl.429S.

[16] W. B. Kannel, R. D. Abbott, D. D. Savage, and P. M. Mcnamara, "Epidemiologic Features of Chronic Atrial Fibrillation: The Framingham Study," *N. Engl. J. Med.*, vol. 306, no. 17, pp. 1018–1022, Apr. 1982, doi: 10.1056/NEJM198204293061703.

[17] A. D. Krahn, J. Manfreda, R. B. Tate, F. A. L. Mathewson, and T. E. Cuddy, "The natural history of atrial fibrillation: Incidence, risk factors, and prognosis in the manitoba follow-up study," *Am. J. Med.*, vol. 98, no. 5, pp. 476–484, 1995, doi: 10.1016/S0002-9343(99)80348-9.

[18] S. B. Olsson *et al.*, "Spontaneous Reversion from Long-Lasting Atrial Fibrillation to Sinus Rhythm," *Acta Med. Scand.*, vol. 207, no. 1–6, pp. 5–20, 1980, doi: 10.1111/j.0954-6820.1980.tb09668.x.

[19] P. PETERSEN and J. GODTFREDSEN, "Atrial Fibrillation—a Review of Course and Prognosis," *Acta Med. Scand.*, vol. 216, no. 1, pp. 5–9, 1984, doi: 10.1111/j.0954-6820.1984.tb03763.x.

[20] E. J. Benjamin, D. Levy, S. M. Vaziri, R. B. D'agostino, A. J. Belanger, and P. A. Wolf, "Independent Risk Factors for Atrial Fibrillation in a Population-Based Cohort: The Framingham Heart Study," *JAMA J. Am. Med. Assoc.*, vol. 271, no. 11, pp. 840–844, Mar. 1994, doi: 10.1001/jama.1994.03510350050036.

[21] V. Fuster *et al.*, "ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation - Executive summary: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Com," *Circulation*, vol. 114, no. 7, pp. 700–752, Aug. 2006, doi: 10.1161/CIRCULATIONAHA.106.177031.

[22] C. Gutierrez and D. G. Blanchard, "Atrial Fibrillation: Diagnosis and Treatment," 2011.

[23] C. R. Kerr *et al.*, "Progression to chronic atrial fibrillation after the initial diagnosis of paroxysmal atrial fibrillation: Results from the Canadian Registry of Atrial Fibrillation," *Am. Heart J.*, vol. 149, no. 3, pp. 489–496, Mar. 2005, doi: 10.1016/j.ahj.2004.09.053.

[24] M. Allessie, J. Ausma, and U. Schotten, "Electrical, contractile and structural remodeling during atrial fibrillation," *Cardiovasc. Res.*, vol. 54, no. 2, pp. 230–246, 2002, doi: 10.1016/S0008-6363(02)00258-4.

[25] C. Aimé-Sempé *et al.*, "Myocardial cell death in fibrillating and dilated human right atria," *J. Am. Coll. Cardiol.*, vol. 34, no. 5, pp. 1577–1586, Nov. 1999, doi:

10.1016/S0735-1097(99)00382-4.

[26] Mayoclinic, “Ventricular tachycardia - Symptoms and causes - Mayo Clinic,” *Mayo Clinic*, 2018. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/ventricular-tachycardia/symptoms-causes/syc-20355138. [Accessed: 12-Jun-2020].

[27] H. L. Greene, P. R. Reid, and A. H. Schaeffer, “The Repetitive Ventricular Response in Man: A Predictor of Sudden Death,” *N. Engl. J. Med.*, vol. 299, no. 14, pp. 729–734, Oct. 1978, doi: 10.1056/NEJM197810052991401.

[28] B. Rodríguez, N. Trayanova, and D. Noble, “Modeling cardiac ischemia,” in *Annals of the New York Academy of Sciences*, 2006, vol. 1080, pp. 395–414, doi: 10.1196/annals.1380.029.

[29] M. Clinic, “Cardiomyopathy - Symptoms and causes - Mayo Clinic,” *Mayo Clinic*, 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/cardiomyopathy/symptoms-causes/syc-20370709. [Accessed: 12-Jun-2020].

[30] J. Lodder, “Cardiac embolism,” *Neurology*, vol. 40, no. 8. p. 1322, 1990, doi: 10.1212/wnl.40.8.1322-c.

[31] P. Kirchhof *et al.*, “2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS,” *European Heart Journal*, vol. 37, no. 38. Werner Budts, 2016, doi: 10.1093/eurheartj/ehw210.

[32] R. C. Davis *et al.*, “Prevalence of atrial fibrillation in the general population and in high-risk groups: The ECHOES study,” *Europace*, vol. 14, no. 11, pp. 1553–1559, 2012, doi: 10.1093/europace/eus087.

[33] P. Kirchhof *et al.*, “Early and comprehensive management of atrial fibrillation: Executive summary of the proceedings from the 2nd AFNET-EHRA consensus conference ‘research perspectives in AF,’” *European Heart Journal*, vol. 30, no. 24. pp. 2969–2980, 2009, doi: 10.1093/eurheartj/ehp235.

[34] J. Engdahl, L. Andersson, M. Mirskaya, and M. Rosenqvist, “Stepwise screening of atrial fibrillation in a 75-year-old population: Implications for stroke prevention,” *Circulation*, vol. 127, no. 8, pp. 930–937, Feb. 2013, doi: 10.1161/CIRCULATIONAHA.112.126656.

[35] G. Kaleschke *et al.*, “Prospective, multicentre validation of a simple, patient-operated electrocardiographic system for the detection of arrhythmias and electrocardiographic changes,” *Europace*, vol. 11, no. 10, pp. 1362–1368, 2009, doi: 10.1093/europace/eup262.

[36] R. G. Tieleman *et al.*, “Validation and clinical use of a novel diagnostic device for screening of atrial fibrillation,” *Europace*, vol. 16, no. 9, pp. 1291–1295, 2014, doi: 10.1093/europace/euu057.

[37] P. M. Barrett *et al.*, “Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring,” *Am. J. Med.*, vol. 127, no. 1, pp. 95.e11-95.e17, 2014, doi: 10.1016/j.amjmed.2013.10.003.

[38] F. R. Quinn and D. Gladstone, “Screening for undiagnosed atrial fibrillation in the community,” *Current Opinion in Cardiology*, vol. 29, no. 1. Curr Opin Cardiol, pp. 28–

35, Jan-2014, doi: 10.1097/HCO.0000000000000018.

[39] J. S. Healey *et al.*, “Subclinical atrial fibrillation and the risk of stroke,” *N. Engl. J. Med.*, vol. 366, no. 2, pp. 120–129, Jan. 2012, doi: 10.1056/NEJMoa1105575.

[40] L. A. Sposato, L. E. Cipriano, G. Saposnik, E. R. Vargas, P. M. Riccio, and V. Hachinski, “Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: A systematic review and meta-analysis,” *Lancet Neurol.*, vol. 14, no. 4, pp. 377–387, Apr. 2015, doi: 10.1016/S1474-4422(15)70027-X.

[41] V. N. Thijs *et al.*, “Predictors for atrial fibrillation detection after cryptogenic stroke: Results from CRYSTAL AF,” *Neurology*, vol. 86, no. 3, pp. 261–269, Jan. 2016, doi: 10.1212/WNL.0000000000002282.

[42] B. B. Love and B. H. Bendixen, “Classification of subtype of acute ischemic stroke definitions for use in a multicenter clinical trial,” *Stroke*, vol. 24, no. 1, pp. 35–41, 1993, doi: 10.1161/01.STR.24.1.35.

[43] S. Higgins, Julian PT and Green, *Cochrane Handbook for Systematic Reviews of Interventions*. 2019.

[44] A. Liberati *et al.*, “The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration,” *PLoS Medicine*, vol. 6, no. 7. p. e1000100, 21-Jul-2009, doi: 10.1371/journal.pmed.1000100.

[45] D. Moher *et al.*, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement,” *Annals of Internal Medicine*, vol. 151, no. 4. American College of Physicians, pp. 264–269, 2009, doi: 10.7326/0003-4819-151-4-200908180-00135.

[46] F. Weidt and R. Silva, “Systematic Literature Review in Computer Science-A Practical Guide,” *Relatórios Técnicos do DCC/UFJF*, vol. 1, no. 0, pp. 1–7, 2016, doi: 10.1027/1016-9040.11.3.244.

[47] “The best research databases for computer science [Update 2019] - Paperpile.” [Online]. Available: https://paperpile.com/g/research-databases-computer-science/. [Accessed: 12-Apr-2020].

[48] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, “Rayyan-a web and mobile app for systematic reviews,” *Syst. Rev.*, vol. 5, no. 1, p. 210, Dec. 2016, doi: 10.1186/s13643-016-0384-4.

[49] “5: Inclusion and exclusion criteria - Systematic Reviews - Library Guides at University of Melbourne.” [Online]. Available: https://unimelb.libguides.com/c.php?g=492361&p=3368110. [Accessed: 15-Jan-2020].

[50] H. ElMoaqet, Z. Almuwaqat, M. Ryalat, and N. Almtireen, “A new algorithm for short term prediction of persistent atrial fibrillation,” *2017 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol.*, pp. 1–6, 2017.

[51] E. Ebrahimzadeh *et al.*, “Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal,” *Comput. Methods Programs Biomed.*, vol. 165, pp. 53–67, 2018.

[52] M. Mohebbi and H. Ghassemian, “Prediction of paroxysmal atrial fibrillation based on

non-linear analysis and spectrum and bispectrum features of the heart rate variability signal," *Comput. Methods Programs Biomed.*, vol. 105, no. 1, pp. 40–49, 2012.

[53] K. H. Boon, M. Khalil-Hani, M. B. Malarvili, and C. W. Sia, "Paroxysmal atrial fibrillation prediction method with shorter HRV sequences," *Comput. Methods Programs Biomed.*, vol. 134, pp. 187–196, 2016.

[54] R. Spronken-Smith, "SJR : Scientific Journal Rankings," *Assessment & Evaluation in Higher Education*, 2012. [Online]. Available: https://www.scimagojr.com/journalrank.php?country=PT. [Accessed: 18-Jan-2020].

[55] B. Schauerte, "Conference Ranks." 2017.

[56] "Scopus - Document search | Signed in." [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 15-Jan-2020].

[57] A. Frank and A. Asuncion, "{UCI} Machine Learning Repository." 2010.

[58] Mayo Clinic, "Mayo Clinic Mayo Clinic." 2019.

[59] M. Laboratory For Computational Physiology, "The MIMIC III Clinical Database," 2015, doi: 10.13026/C2XW26.

[60] G. B. Moody, A. L. Goldberger, S. McClennen, and S. P. Swiryn, "Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001," *Comput. Cardiol.*, pp. 113–116, 2001, doi: 10.1109/CIC.2001.977604.

[61] R. Questionnaire, "China Kadoorie Biobank," 2013. [Online]. Available: https://www.ckbiobank.org/site/. [Accessed: 18-Jan-2020].

[62] Y. Shen, Y. Yang, S. Parish, Z. Chen, R. Clarke, and D. A. Clifton, "Risk prediction for cardiovascular disease using ECG data in the China kadoorie biobank," *2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 2419–2422, 2016.

[63] K. H. Boon, M. Khalil-Hani, and M. B. Malarvili, "Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III," *Comput. Methods Programs Biomed.*, vol. 153, pp. 171–184, 2018.

[64] H. Costin, C. Rotariu, A. Păsărică, and ru, "Atrial fibrillation onset prediction using variability of ECG signals," *2013 8th Int. Symp. Adv. Top. Electr. Eng.*, pp. 1–4, 2013.

[65] M. Mohamed, M. Farah, and A. Fahed, "Evidence-based prediction of Atrial Fibrillation using physiological signals," *2019 3rd Int. Conf. Bio-engineering Smart Technol.*, pp. 1–4, 2019.

[66] J. Kim, O. Sangjun, Y. Kim, and M. Lee, "Convolutional Neural Network with Biologically Inspired Retinal Structure," *7th Annu. Int. Conf. Biol. Inspired Cogn. Archit. BICA 2016, held July 16 to July 19, 2016 New York City, NY, USA*, vol. 88, pp. 145–154, 2016.

[67] Z. I. Attia *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.

[68] S. Rajalakshmi and K. V Madhav, "A collaborative prediction of presence of Arrhythmia in human heart with electrocardiogram data using machine learning algorithms with analytics," *J. Comput. Sci.*, vol. 15, no. 2, pp. 278–287, 2019.

[69] Z. Li, H. Derksen, J. Gryak, H. Ghanbari, P. Gunaratne, and K. Najarian, "A Novel Atrial

Fibrillation Prediction Algorithm Applicable to Recordings from Portable Devices," *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4034–4037, 2018.

[70] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985, doi: 10.1109/TBME.1985.325532.

[71] MATLAB, "MATLAB documentation," *MATLAB*. p. R2012b, 2012, doi: 10.1201/9781420034950.

[72] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 2014, pp. 675–678, doi: 10.1145/2647868.2654889.

[73] C. C. Chang and C. J. Lin, "LIBSVM: A Library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, doi: 10.1145/1961189.1961199.

[74] International Organization for Standardization, "International Standard ISO/IEC 14882:2014(E) – Programming Language C++," p. 1358, 2014.

[75] et al. Abadi, Mart'in, Barham P, Chen J, Chen Z, Davis A, Dean J, "Tensorflow: A system for large-scale machine learning," in *12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16)*, 2016, pp. 265–83.

[76] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of Human Electrocardiogram for Biometric Recognition," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, pp. 1–11, Sep. 2007, doi: 10.1155/2008/148658.

[77] D. M. German, M. M. Kabir, T. A. Dewland, C. A. Henrikson, and L. G. Tereshchenko, "Atrial Fibrillation Predictors: Importance of the Electrocardiogram," *Annals of Noninvasive Electrocardiology*, vol. 21, no. 1. Blackwell Publishing Inc., pp. 20–29, 01-Jan-2016, doi: 10.1111/anec.12321.

[78] S. Cheng *et al.*, "Long-term outcomes in individuals with prolonged PR interval or first-degree atrioventricular block," *JAMA - J. Am. Med. Assoc.*, vol. 301, no. 24, pp. 2571–2577, Jun. 2009, doi: 10.1001/jama.2009.888.

[79] J. W. Magnani, M. A. Williamson, P. T. Ellinor, K. M. Monahan, and E. J. Benjamin, "P wave indices current status and future directions in epidemiology, clinical, and research applications," *Circulation: Arrhythmia and Electrophysiology*, vol. 2, no. 1. Circ Arrhythm Electrophysiol, pp. 72–79, Feb-2009, doi: 10.1161/CIRCEP.108.806828.

[80] J. S. Steinberg, S. Zelenkofske, S. C. Wong, M. Gelernt, R. Sciacca, and E. Menchavez, "Value of the P-wave signal-averaged ECG for predicting atrial fibrillation after cardiac surgery," *Circulation*, vol. 88, no. 6, pp. 2618–2622, 1993, doi: 10.1161/01.CIR.88.6.2618.

[81] M. Budeus, M. Hennersdorf, C. Perings, H. Wieneke, R. Erbel, and S. Sack, "Prediction of the recurrence of atrial fibrillation after successful cardioversion with P wave signal-averaged ECG," *Ann. Noninvasive Electrocardiol.*, vol. 10, no. 4, pp. 414–419, Oct. 2005, doi: 10.1111/j.1542-474X.2005.00059.x.

[82] J. B. Nielsen *et al.*, "P-wave duration and the risk of atrial fibrillation: Results from the Copenhagen ECG Study," *Hear. Rhythm*, vol. 12, no. 9, pp. 1887–1895, Sep. 2015, doi: 10.1016/j.hrthm.2015.04.026.

[83] J. J. MORRIS, E. H. ESTES, R. E. WHALEN, H. K. THOMPSON, and H. D. MCINTOSH, "P-WAVE ANALYSIS IN VALVULAR HEART DISEASE.," *Circulation*, vol. 29, no. 2, pp. 242–252, Feb. 1964, doi: 10.1161/01.CIR.29.2.242.

[84] S. Kohsaka, R. R. Sciacca, K. Sugioka, R. L. Sacco, S. Homma, and M. R. Di Tullio, "Electrocardiographic left atrial abnormalities and risk of ischemic stroke," *Stroke*, vol. 36, no. 11, pp. 2481–2483, Nov. 2005, doi: 10.1161/01.STR.0000185682.09981.26.

[85] E. Z. Soliman, R. J. Prineas, L. D. Case, Z. M. Zhang, and D. C. Goff, "Ethnic distribution of ecg predictors of atrial fibrillation and its impact on understanding the ethnic distribution of ischemic stroke in the atherosclerosis risk in communities (ARIC) study," *Stroke*, vol. 40, no. 4, pp. 1204–1211, Apr. 2009, doi: 10.1161/STROKEAHA.108.534735.

[86] M. V. Perez *et al.*, "Electrocardiographic predictors of atrial fibrillation," *Am. Heart J.*, vol. 158, no. 4, pp. 622–628, Oct. 2009, doi: 10.1016/j.ahj.2009.08.002.

[87] E. Tülümen *et al.*, "PQ segment depression in patients with short QT syndrome: A novel marker for diagnosing short QT syndrome?," *Hear. Rhythm*, vol. 11, no. 6, pp. 1024–1030, 2014, doi: 10.1016/j.hrthm.2014.02.024.

[88] Y. Zhang, W. S. Post, E. Blasco-Colmenares, D. Dalal, G. F. Tomaselli, and E. Guallara, "Electrocardiographic QT interval and mortality: A meta-analysis," *Epidemiology*, vol. 22, no. 5, pp. 660–670, Sep. 2011, doi: 10.1097/EDE.0b013e318225768b.

[89] C. Giustetto *et al.*, "Long-term follow-up of patients with short QT syndrome," *J. Am. Coll. Cardiol.*, vol. 58, no. 6, pp. 587–595, Aug. 2011, doi: 10.1016/j.jacc.2011.03.038.

[90] J. N. Johnson, D. J. Tester, J. Perry, B. A. Salisbury, C. R. Reed, and M. J. Ackerman, "Prevalence of early-onset atrial fibrillation in congenital long QT syndrome," *Hear. Rhythm*, vol. 5, no. 5, pp. 704–709, May 2008, doi: 10.1016/j.hrthm.2008.02.007.

[91] J. Chrispin *et al.*, "Association of electrocardiographic and imaging surrogates of left ventricular hypertrophy with incident atrial fibrillation: MESA (Multi-Ethnic Study of Atherosclerosis)," *J. Am. Coll. Cardiol.*, vol. 63, no. 19, pp. 2007–2013, May 2014, doi: 10.1016/j.jacc.2014.01.066.

[92] S. K. Agarwal, G. Heiss, P. M. Rautaharju, E. Shahar, M. W. Massing, and R. J. Simpson, "Premature ventricular complexes and the risk of incident stroke: The atherosclerosis risk in communities (ARIC) study," *Stroke*, vol. 41, no. 4, pp. 588–593, Apr. 2010, doi: 10.1161/STROKEAHA.109.567800.

[93] T. To, "Catheter Ablation for Ectopic Heartbeats."

[94] M. Haïssaguerre *et al.*, "Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins," *N. Engl. J. Med.*, vol. 339, no. 10, pp. 659–666, Sep. 1998, doi: 10.1056/NEJM199809033391003.

[95] J. O. Smith, "Mathematics of the Discrete Fourier Transform," *Identity*, p. 322, 2007.

[96] A. LEOPA, "THEORETICAL ANALYSIS OF DYNAMIC RESPONSE OF A BRIDGE SECTION LOADING BY THE IMPULSIVE FORCE," *J. Eng. Stud. Res.*, vol. 18, no. 2, Mar. 2017, doi: 10.29081/jesr.v18i2.218.

[97] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi:

10.1109/5.726791.
[98] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.
[99] V. Jain and H. S. Seung, “Natural image denoising with convolutional networks,” in *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, 2009, pp. 769–776.
[100] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Mitosis detection in breast cancer histology images with deep neural networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8150 LNCS, no. PART 2, pp. 411–418, doi: 10.1007/978-3-642-40763-5_51.
[101] S. C. Turaga *et al.*, “Convolutional networks can learn to generate affinity graphs for image segmentation,” *Neural Computation*, vol. 22, no. 2. Neural Comput, pp. 511–538, Feb-2010, doi: 10.1162/neco.2009.10-08-881.
[102] R. Li, “Machine Learning Methods for Medical and Biological Image Computing,” in *Climate Change 2013 - The Physical Science Basis*, 2016, pp. 1–30.
[103] I. G. and Y. Bengio and A. Courville, *Deep Learning*. MIT Press, 2016.
[104] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.,” *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.
[105] Python Software Foundation, “Python 3.0 Release | Python.org,” 2009. [Online]. Available: https://www.python.org/download/releases/3.0/. [Accessed: 23-May-2020].
[106] “wfdb · PyPI.” [Online]. Available: https://pypi.org/project/wfdb/. [Accessed: 08-Jun-2020].
[107] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[108] G. Ellis, “Filters in Control Systems,” in *Control System Design Guide*, Elsevier, 2012, pp. 165–183.
[109] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 99–104, May 2007, doi: 10.1109/MCSE.2007.55.
[110] Chollet François, “Keras: The Python Deep Learning library,” *Keras.Io*, 2015, doi: 10.1086/316861.
[111] P. Lameski, E. Zdravevski, S. Koceski, A. Kulakov, and V. Trajkovik, “Suppression of Intensive Care Unit False Alarms Based on the Arterial Blood Pressure Signal,” *IEEE Access*, vol. 5, pp. 5829–5836, 2017, doi: 10.1109/ACCESS.2017.2690380.
[112] “AF Termination Challenge Database v1.0.0.” [Online]. Available: https://physionet.org/content/afpdb/1.0.0/. [Accessed: 08-Jun-2020].

# Appendix A - Articles

## “Prediction of Atrial Fibrillation using Artificial Intelligence on Electrocardiograms: A Systematic Review”

Submitted

# Prediction of Atrial Fibrillation using Artificial Intelligence on Electrocardiograms: A Systematic Review

Igor Matias[a,b,16], Nuno Garcia[a,b], Sandeep Pirbhulal[a,c], Virginie Felizardo[a,b], Nuno Pombo[a,b], Henriques Zacarias[a,b], Miguel Sousa[a], Eftim Zdravevski[d,e]

[a]*Universidade da Beira Interior, Covilhã, Portugal*

[b]*Instituto de Telecomunicações, Covilhã, Portugal*

[c]*Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway*

[d]*Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, Skopje, North Macedonia*

[e]*COPELABS, Universidade Lusófona de Humanidades e Tecnologias, Lisboa, Portugal*

**Abstract** – Atrial Fibrillation (AF) is a type of Cardiovascular Disease (CVD) characterized by irregular heartbeats, with four different types, two of which are complicated to diagnose using standard techniques such as the Electrocardiogram (ECG). However, and because smart wearables are increasingly a piece of commodity equipment, there are several ways of detecting and predicting AF episodes using only an ECG exam, allowing physicians to do an easier diagnose of this condition. By searching in several databases, this study presents a review of all the articles published in the last ten years, focusing only on those who reported studies using Artificial Intelligence (AI) for prediction of AF, on the six databases that were selected. The results show that only twelve studies were selected for this systematic review, where three of them applied deep learning techniques (25%), six of them used machine learning methods (50%) and three others focused on applying general artificial intelligence models (25%), such as statistical methods. In conclusion, this study revealed that the prediction of AF is yet an under-developed field in the context of AI, and deep learning techniques are increasing the accuracy, but these are not as frequently applied as it would be expected. Also, more than half of the selected studies were published since the year of 2016, allowing the conclusion that this topic is very recent and probably has a high potential for additional research.

**Keywords** – ECG waveform, Electrocardiogram, Artificial Intelligence, prediction algorithms, Atrial Fibrillation

## 1. INTRODUCTION

The World Health Organization (WHO) states that Cardiovascular Diseases (CVDs) are associated with heart and blood vessels-based disorders. CVDs may include

[16] Corresponding author.
Email address: igor.matias@ubi.pt.
Postal address: Universidade da Beira Interior, Departamento de Informática, Rua Marquês d'Ávila e Bolama, 6201-001, Covilhã, Portugal.

hypertension, heart attack, cerebrovascular disease, heart failure, rheumatic congenital heart disease, and cardiomyopathies, among others [1].
CVDs were the most significant cause of death worldwide in 2016 [1], adding to approximately 27% of all the estimated deaths in the world [2] and about 45% in Europe by 2017 [3]. Solely in the European Union, in the year of 2006, CVDs were estimated to have cost €104 thousand million in healthcare [4].
Among the CVDs, Atrial Fibrillation (AF) is an arrhythmia, which is characterized by irregular heartbeats, that can lead to blood clots, heart failure, stroke and other heart-related complications including death, and is commonly underdiagnosed [5], [6]. It can assume four different types, Paroxysmal AF, Persistent AF, Long-standing Persistent AF and Permanent AF. According to [7], only the severest Long-standing Persistent and Permanent can be easily detected with an ECG exam, the other types being harder to identify due to the irregularity of its symptoms.
This irregularity of symptoms makes it very hard to detect both less severe types of AF, due to the high probability of these patients not presenting any symptoms during an ECG. To avoid this low efficiency of detection of AF, predictive models were being developed, allowing the diagnosis of a patient's AF state only based in a short ECG signal, avoiding extra-long and intrusive devices and methodologies.
The detection of AF is commonly performed by analysing the signal collected from an ECG, a non-invasive and painless exam with quick results, typically outputting several charts resulting from a 12 lead collection setup [8].
Nowadays, portable devices such as smartwatches, smart fitness bands or portable medical signal collectors have a crucial role in evolving the way we diagnose several health disorders before they step into a high-risk medical field. It is due to its ease on the recording, for example, ECG and pulse signals [9]–[12], being always present with the patient itself, being thus able to collect data from several moments of the day, within different activity and emotional states. Some of these devices, despite using a smaller number of ECG leads, sometimes 3 or 2, have been proved to be as efficient as Hospital grade ECG equipment, as tested in [13].
However, in the last years, there were developed several new methods to detect and to predict the existence of the different types of AF. These new approaches all require powerful algorithms combined with innovative sensors, applying several different types of AI.
There is a multitude of benefits from integrating AI into healthcare, including automation tasks and analysing big patient data sets to deliver better healthcare faster, and at a lower cost [14]. The usage of AI into healthcare, and consequently AF detection and prediction, does allow the analysis of bigger datasets, with the faster result, easing the workload of healthcare professionals, making possibly automated and real-time diagnosis, anytime and anywhere.
This paper presents a systematic literature review on ECG-based models for AF Prediction using AI techniques covering the last ten years. At the time of this review, we did not find any report that covers this topic. Therefore, the selected studies reviewed here present the most recent work in this field.

The main contributions of this article are:
1) present a discussion on how the prediction of AF has been and is currently addressed;
2) indicates what databases, features, pre-processing and predictive algorithms have been and are presently used in these systems;
3) a benchmark to conclude which achieved models performs better.

The remainder of this paper is organized as follows: Section 2 presents a description of the method that was designed for eligibility selection and extraction of information. Section 3 includes the results of the search by displaying the selected studies and their features in summary tables. The discussion and the answer to the research questions are presented in Section 4. Finally, Section 5 of this review shows the highlights and limitations of this study.

## 2. METHODS

This systematic literature review was conducted informed by recommendations from the Cochrane Handbook for Systematic Reviews of Interventions, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [15]–[17], and based on the guidelines from [18].
This section explains in detail the methodology used for conducting this review.

### 2.1. SEARCH STRATEGY

The databases Web of Science[17], Scopus[18], ACM Digital Library[19], IEEE Xplore[20], PubMed[21] and Science Direct[22] were used to search for relevant peer-reviewed publications from January 1, 2009, 00:00 to December 13, 2019, 04:22.
The first two used databases are interdisciplinary databases. ACM Digital Library is, according to [19], the number one database related to academic databases for computer science and IEEE Xplore was chosen due to its high number of articles from the field of computer science. Finally, PubMed was used due to its content regarding research in biomedicine and Science Direct because of its high number of articles from thousands of books and journals.
We searched titles and abstracts using the keywords presented below. The list of references from the selected articles was manually screened for the inclusion of additional relevant articles.
The keywords used in all the databases were:

> (“machine learning” OR “artificial intelligence”) AND (“ECG” OR electrocardio*) AND (“Atrial Fibrillation” OR “AF” OR “arrhythmia”) AND ("prediction" OR "prognosis" OR "foresee").

### 2.2. STUDY SELECTION

We screened the titles and abstracts of all identified publications for eligibility, using the web application Rayyan QCRI [20].

---

[17] S. C. Collection, “Web of Science [v.5.14] - Web of Science Core Collection引用レポート,” pp. 8–9
[18] “Scopus - Document search | Signed in.” [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 15-Jan-2020]
[19] C. The et al., “ACM Digital Library,” 1985
[20] “IEEE Xplore Digital Library.” [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: 15-Jan-2020]
[21] “Home - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/. [Accessed: 15-Jan-2020]
[22] "ScienceDirect.com | Science, health and medical journals, full-text articles and books." [Online]. Available: https://www.sciencedirect.com/. [Accessed: 15-Jan-2020]

The inclusion criteria were broadly defined to increase the sensitivity of the search. The aim was to identify the articles that applied any AI method on ECG signals for prediction of AF on patients with no previous clinical conditionings.
Additional inclusion/exclusion criteria are summarized in Table 1.

*Table 1 - Inclusion and exclusion criteria used in the review, as in* [21].

| **Type** | **Inclusion** | **Exclusion** |
| --- | --- | --- |
| Date | All | None |
| Exposure of interest | All | None |
| Geographic location of study | All | None |
| Language | English | Any other language |
| Participants | With no recent surgical procedures or drugs effects during the ECG collection | With any recent surgical procedure or ingestion or drugs effects during the ECG collection |
| Peer review | Journal and Conference | All others |
| Reported outcomes | At least one: accuracy, sensitivity, specificity, confusion matrix | All others that did not report any metric |
| Setting | All | None |
| Study design | All | None |
| Type of publication | Journal and Conference | All others |

According to the inclusion and exclusion criteria presented in Table 1, all the articles not excluded after its analysis had its full texts reviewed for eligibility.

### 2.3. EXTRACTION OF STUDY CHARACTERISTICS

The extraction of information from the selected publications was based on the pre-defined categories, to collect the relevant data and to assess, analyse the model characteristics and its experimental setup:

- Study Information: defines the study citation and year of publication;
- Inputs: assess the inputs used to develop the algorithm, including dataset used and amount and age of the individuals from where the dataset was collected;
- Signal treatment: defines the usage of the ECG signals received as input, namely the features extracted from it, the duration of the signal used for training, and the tools used for the process;
- Methods: defines the methods/algorithms applied to the pre-processing of the ECG signal, the prediction of AF and evaluation of the model, as well as the number of iterations, and the data split applied for training and testing;
- Performances: defines the evaluation metrics used to assess the predictions.

### 2.4. RESEARCH QUESTIONS

The research questions of this review were:

- (RQ1) How is the prediction problem assessed?
- (RQ2) What databases and features are used?
- (RQ3) What pre-processing algorithms are used?

- (RQ4) What predictive algorithms are used?
- (RQ5) Which are the models that perform the best?

The (RQ1) motivation was to identify the trends and possible opportunities for research topic focus.
The motivation for (RQ2) and (RQ3) was to identify new advances on features and databases and pre-processing techniques used for prediction of AF, respectively.
The motivation for (RQ4) was to identify the new predictive algorithms used to predict AF using ECG data on recent studies.
Finally, for (RQ5) motivation, it was intended to identify the models that can more accurately predict AF episodes, this way identifying trends and possible opportunities for the use of research methods.

## 3. RESULTS

At the beginning of the search, it yielded 375 unique records, after the removal of duplicates.
After the review of the title and abstract and following the inclusion and exclusion criteria presented in Table 1, 293 records were excluded; 82 full-text publications were assessed for eligibility and after full-text review, of which 72 records were excluded (Figure 1).

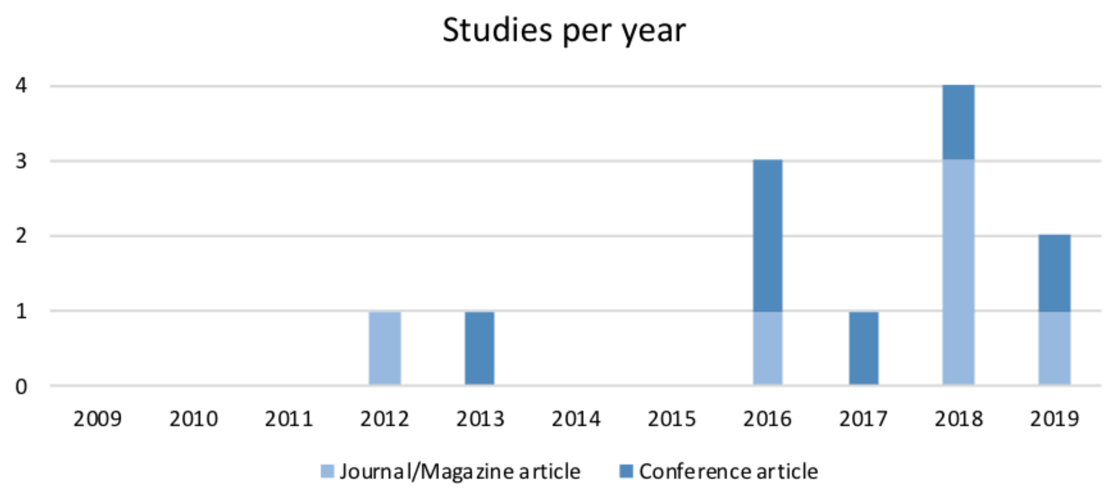



*Figure 1 - Number of studies from 2009 to 2019.*

The excluded records can be described as follows. Sixty-four studies reported research related to AF, but there was no prediction of AF during its execution. Two studies could not be fully read because the authors of this systematic literature review were not able to obtain the full articles. Two articles did not present the evaluation metrics included in the Inclusion Criteria of this search presented in Table 1. Two studies were focused on reviewing state-of-the-art related to AF identification. One study had a publication date before 2009, and another one did the work with ECG collected from patients with surgical proceedings (prophylactic ICD-implantation).
From the remaining 10 records, reference tracking was performed, and two studies were added, totalizing 12 studies to be included for the data extraction and the qualitative synthesis stage. The flow diagram of the identification and inclusion of articles is shown in Figure 2.

### 3.1. ELIGIBILITY OF THE STUDIES

Despite all the selected studies that met the inclusion and exclusion criteria, it is useful to clarify the selection of some studies.

The study [22] presents an algorithm for short term prediction of Persistent AF, but the ECG data used was collected from sheep instead of human individuals. Despite this, we considered that this article is eligible, not so much because of the nature of the ECG signal, but mostly because of the described methodologies and algorithmic approaches the paper describes.

In the studies [23], [24] and [25], the prediction of AF was only performed between pre and post AF moments, not allowing for cases with no AF prediction. However, they were included because of the insight the papers report to this research.

Finally, in the study [25], the reported measurements with a single fold method match neither the tables nor the text of the paper. It was decided to include this last study, but only to consider the best measurements for the 10-fold method, that has valid reporting of values in the tables and the study's text.

### 3.2. SOURCE OF EVIDENCE

To be able to evaluate and classify the selected studies, Table 2, Table 3, Table 4, Table 5 and Table 6 present statistical data about the publication year, the ranking of the magazine the article was published in, the type of publication, the geographic region the study covered and the number of citations the article has.

*Table 2 - Number of publications by year of publication.*

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| Publication year | 2009 – 2016 | 5 | 41,67% |
| | 2017 – 2018 | 5 | 41,67% |
| | 2019 | 2 | 16,67% |

*Table 3 - Ranking of the article's publication place, according to* [26] *and* [27].

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| Ranking of article's publication place, accordingly to [26] and [27] | 1st quarter Journal | 5 | 41,67% |
| | 4th quarter Journal | 1 | 8,33% |
| | H-factor = 83 Journal | 4 | 33,33% |
| | H-factor = 23 Journal | 1 | 8,33% |
| | H-factor = 700 Journal | 1 | 8,33% |
| | 0 <= H-factor <= 10 Conference Proceeding | 3 | 25,00% |
| | 11 <= H-factor <= 20 Conference Proceeding | 2 | 16,67% |
| | B1 ranked Conference | 2 | 16,67% |
| | B4 ranked Conference | 1 | 8,33% |

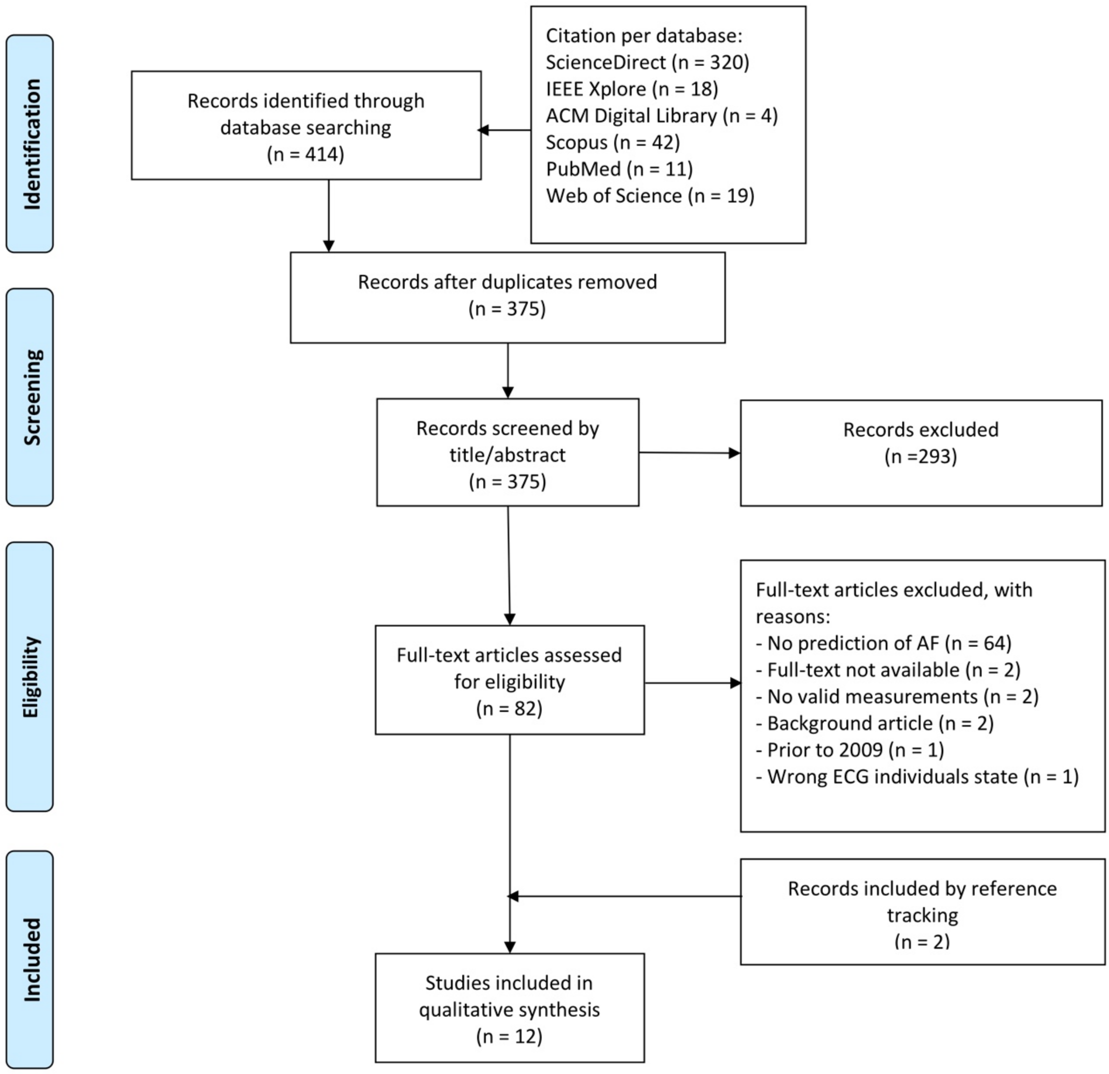



*Figure 2 - Flow diagram of identification and inclusion papers.*

*Table 4 - Number of publications by journal or conference type.*

| | **Description** | **Number of studies** | **Portion of total** |
|---|---|---|---|
| Journal or Conference type | Medicine Journal | 1 | 8,33% |
| | Bioinformatics Journal | 4 | 33,33% |
| | Computer Science Journal | 1 | 8,33% |
| | IEEE Conference or sponsored by IEEE | 3 | 25,00% |

*Table 5 - Continent where the studies were conducted.*

| | **Description** | **Number of studies** | **Portion of total** |
|---|---|---|---|
| The continent where it was conducted | Africa | 1 | 8,33% |
| | Asia | 6 | 50,00% |
| | Europe | 3 | 25,00% |
| | America | 2 | 16,67% |

*Table 6 - Number of citations per selected articles, according to* [28].

| | Description | Number of studies | Portion of total |
|---|---|---|---|
| Publication citations, according to [17] | 0 | 4 | 33,33% |
| | 2 | 1 | 8,33% |
| | 4 | 1 | 8,33% |
| | 5 | 1 | 8,33% |
| | 7 | 1 | 8,33% |
| | 8 | 1 | 8,33% |
| | 12 | 1 | 8,33% |
| | 15 | 1 | 8,33% |
| | 48 | 1 | 8,33% |

### 3.3. STUDY PARTICIPANTS AND DESIGN

Seven studies (58,33%) were based on databases with small samples of individuals (less than 100), two studies (16,67%) with samples between 100 and 25000 individuals, one study (8,33%) with a sample of around 126000 individuals, and two studies (16,67%) did not report the sample size.
Three studies (25,00%) used a personal database of ECG records, one (8,33%) used a UCI Repository Warehouse's ([29]) database, one was based on the Mayo Clinic ([30]) ECG Laboratory's database, one used the Medical Information Mart for Intensive Care III database ([31]), five articles (41,67%) have done the research using the Atrial Fibrillation Prediction Database ([32]), and one study (8,33%) used a China Kadoorie Biobank's ([33]) database. Both [31] and [32] datasets are available at the Physionet Repository. Only [31], [32] and [33] are publicly available.

### 3.4. PREDICTION METHODS

The selected articles used several different methods of AI for prediction of AF:

- Five articles ([22], [24], [25], [34], [35]) (41,67%) applied Support Vector Machine;
- Two articles ([36], [37]) (16,67%) used statistical AI methods;
- Two articles ([38], [39]) (16,67%) used Convolutional Neural Network;
- One study ([40]) (8,33%) applied its study using Arrhythmia Fuzzy Hybrid Classifier;
- Another study ([41]) (8,33%) used Markov Chain;
- Finally, the last of all articles ([23]) used the method Mixture of Experts for prediction of AF.

### 3.5. DATA COLLECTED FROM SELECTED STUDIES

During the quality synthesis process, it is essential to get as much information from the selected studies as possible. However, despite all the articles having some extra data, some of it was not comparable, the reason why they are not mentioned in the next collected data tables.
Table 7 shows the dataset used in each one of the selected studies, including the number of individuals, where the data came from, and its age, if provided.

*Table 7 - Input information collected from studies. NR=Not Reported/Not Applicable.*

| Year | Study | Dataset used | Number of participants | Age of participants |
|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [24] | Atrial Fibrillation Prediction Database [32] | NR | NR |
| 2013 | Costin *et al.* [36] | Atrial Fibrillation Prediction Database [32] | 75 | NR |
| 2016 | Kim *et al.*[38] | Own collected dataset | 1 | NR |
| 2016 | Shen *et al.*[34] | China Kadoorie Biobank [33] | 24369 | NR |
| 2016 | Boon *et al.*[25] | Atrial Fibrillation Prediction Database [32] | 53 | NR |
| 2017 | ElMoaqet *et al.*[22] | Own collected dataset | 33 | NR |
| 2018 | Rajalakshmi *et al.*[40] | UCI Repository Warehouse [29] | NR | NR |
| 2018 | Li *et al.*[41] | Own collected dataset | 5 | NR |
| 2018 | Boon *et al.*[35] | Atrial Fibrillation Prediction Database [32] | 53 | NR |
| 2018 | Ebrahimzadeh *et al.* [23] | Atrial Fibrillation Prediction Database [32] | 53 | NR |
| 2019 | Attia *et al.*[39] | Mayo Clinic ECG Laboratory [30] | 126526 | >18, average 60,3 |
| 2019 | Mohamed *et al.*[37] | Medical Information Mart for Intensive Care III database [31] | 246 | NR |

From all the twelve selected studies, not all of them apply AI methods that do not need feature selection and extraction from the source ECG signal. Table 8 presents the number of frequency-domain, time-domain, space-domain and non-linear features extracted from each one of the studies, as well as the signal duration used as input to the AI model/method and the tools used at the collecting and pre-processing phase of the studies.

Regarding the signal duration used in each one of the selected studies, it is a noticeable difference between the minimum and maximum among all. The majority used a signal of 300 seconds (three studies), followed by 30 and 10 seconds (two studies each). Some other articles reported usage of signals with 120, 1800 and 3600 seconds length (one study each).

*Table 8 - Signal treatment information collected from studies. NR=Not Reported/Not Applicable.*

| Year | Study | Features extracted from ECG signal | Signal duration (seconds) | Tools used |
|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [24] | 4 frequency-domain<br>6 time-domain<br>4 non-linear | 300 | NR |
| 2013 | Costin *et al.* [36] | 1 frequency-domain<br>1 time-domain | 300 | Pan-Tompkins algorithm [42], MATLAB 2008 [43] |
| 2016 | Kim *et al.*[38] | NR | 30 | Caffe deep learning framework [44] |
| 2016 | Shen *et al.*[34] | 1 time-domain<br>1 space-domain | 10 | NR |
| 2016 | Boon *et al.*[25] | 8 frequency-domain<br>1 time-domain | 1800 | NR |
| 2017 | ElMoaqet *et al.*[22] | 1 frequency-domain<br>5 time-domain<br>3 non-linear | 30 | MATLAB [43], LibSVM toolbox [45] |
| 2018 | Rajalakshmi *et al.*[40] | 5 time-domain | NR | Excel[23], MATLAB 2015 [43], Rapid Miner[24] |
| 2018 | Li *et al.*[41] | NR | 120 | NR |
| 2018 | Boon *et al.*[35] | 3 frequency-domain<br>2 time-domain<br>2 non-linear | 900 | C++ [46], LibSVM library [45] |
| 2018 | Ebrahimzadeh *et al.* [23] | 4 frequency-domain<br>5 time-domain<br>8 non-linear<br>11 time-frequency | 300 | NR |

[23] Microsoft Portugal, "Microsoft Excel," 2019. [Online]. Available: https://products.office.com/pt-pt/excel?legRedir=true&CorrelationId=3e4e9d3a-7d82-42a5-977c-fa3f430fa6ce&rtc=1. [Accessed: 29-Jan-2020]

[24] RapidMiner, "Lightning Fast Data Science Platform for Teams | RapidMiner©," RapidMiner, 2019. [Online]. Available: https://rapidminer.com/. [Accessed: 29-Jan-2020]

| | | | | |
|---|---|---|---|---|
| 2019 | Attia *et al.*[39] | NR | 10 | GE-Marquette ECG machine[25], MUSE system[26], Keras[27], TensorFlow [47], Python[28], R[29] |
| 2019 | Mohamed *et al.*[37] | 5 time-domain | 3600 | NR |

Table 9 has information about the different methods used in each one of the selected studies. It is divided into the methods used for pre-processing the input data for the prediction phase and the performance evaluation. The table also includes the number of iterations used on the training as well as the data split between training and testing subsets.

*Table 9 - Methods applied by the selected studies. NR=Not Reported/Not Applicable.*

| **Year** | **Study** | **Pre-processing method(s)** | **Prediction method(s)** | **Evaluation method(s)** | **Number of iterations** | **Data split (training/ testing) %** |
|---|---|---|---|---|---|---|
| 2012 | Mohebbi *et al.* [24] | Noise removal, QRS detection | Support Vector Machine | NR | NR | 47/53 |
| 2013 | Costin *et al.* [36] | Noise removal | HRV analysis and Morphologic Variability of QRS complexes | NR | NR | 50/50 |
| 2016 | Kim *et al.*[38] | NR | Convolutional Neural Network with | NR | 30000 | 90/10 |

[25] "MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020]
[26] "MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020]
[27] "Home - Keras Documentation." [Online]. Available: https://keras.io/. [Accessed: 29-Jan-2020]
[28] Python Software Foundation, "Welcome to Python.org," 2001, 2019. [Online]. Available: https://www.python.org/. [Accessed: 29-Jan-2020]
[29] The R Foundation, "R: The R Project for Statistical Computing," 2018. [Online]. Available: https://www.r-project.org/. [Accessed: 29-Jan-2020]

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | ON/OFF ReLU | | | |
| 2016 | Shen *et al.*[34] | NR | Support Vector Machine | 5-fold Cross-Validation | NR | NR |
| 2016 | Boon *et al.*[25] | Hamilton and Tompkins algorithm, McNames algorithm | Support Vector Machine | 10-fold Cross-Validation | 10 | 90,6/9,4 |
| 2017 | ElMoaq et *et al.*[22] | Noise removal | Weighted Support Vector Machine | 10-fold Cross-Validation | 100 | 75/25 |
| 2018 | Rajalakshmi *et al.*[40] | Normalisation, Missing values removal | Novel Arrhythmia Fuzzy Hybrid Classifier Algorithm | NR | NR | NR |
| 2018 | Li *et al.*[41] | Noise removal, QRS detection | Markov Chain | NR | NR | NR |
| 2018 | Boon *et al.*[35] | McNames algorithm | Support Vector Machine | 10-fold Cross-Validation | 5 | 90,6/9,4 |
| 2018 | Ebrahimzadeh *et al.* [23] | Noise removal, QRS detection | Mixture of Experts | 10-fold Cross-Validation | NR | 47/53 |
| 2019 | Attia *et al.*[39] | NR | Convolutional Neural Network | NR | NR | 70/20 |
| 2019 | Mohamed *et al.*[37] | NR | Belief Functions Theory | NR | 30 | 67/33 |

The identified models/algorithms in all selected studies were compared with the reported accuracy. Some of them did not report the sensitivity and specificity, neither the F-Score nor the Area Under the Curve. Table 10 contains information about the achievements of each study.

*Table 10 - Evaluation of the selected studies. NR=Not Reported/Not Applicable.*

| **Year** | **Study** | **Accuracy** | **Sensitivity** | **Specificity** | **F-Score** | **Area Under Curve** |
|---|---|---|---|---|---|---|
| 2012 | Mohebbi *et al.* | 96,64% | 96,30% | 93,10% | NR | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | [24] | | | | | |
| 2013 | Costin *et al.* [36] | 90,00% | 89,44% | 89,29% | NR | 89,40% |
| 2016 | Kim *et al.*[38] | 83,58% | NR | NR | NR | NR |
| 2016 | Shen *et al.*[34] | 75,60% | NR | NR | NR | 83,00% |
| 2016 | Boon *et al.*[25] | 80,20% | 81,10% | 79,30% | NR | NR |
| 2017 | ElMoaqet *et al.*[22] | 84,90% | 66,70% | 97,00% | NR | 93,50% |
| 2018 | Rajalakshmi *et al.*[40] | 82,80% | 0,40% | 0,43% | 1,21% | NR |
| 2018 | Li *et al.*[41] | 82,00% | 86,00% | 80,00% | 74,51% | 90,88% |
| 2018 | Boon *et al.*[35] | 87,70% | 86,80% | 88,70% | NR | NR |
| 2018 | Ebrahimzadeh *et al.* [23] | 98,21% | 100,00% | 96,55% | NR | NR |
| 2019 | Attia *et al.*[39] | 83,30% | 82,30% | 83,40% | 45,40% | 90,00% |
| 2019 | Mohamed *et al.*[37] | 70,49% | 77,07% | 63,90% | NR | NR |

As Table 8 indicated, almost all of the selected studies performed feature extraction, with an exception for those who did implement a deep learning method only or did not report this information in the article. Going deeper into the analysis and comparison of the selected studies, Table 11 presents all the features selected and extracted by each one of them, ordering them referring to the number of studies using the same feature, decreasingly.

*Table 11 - Features extracted from input on each one of the selected articles.*

| **Domain** | **Features** | **Studies** |
|---|---|---|
| Frequency | Low-frequency band power (LF) | [24], [23], [35] |
| | High-frequency band power (HF) | [24], [23] |
| | LF/HF ratio | [36], [23] |
| | Low-frequency component of Fast Fourier Transforms (FFT-LF) | [25] |
| | High-frequency component of Fast Fourier Transforms (FFT-HF) | [25] |
| | LL-H1 | [25], [35] |
| | LL-H2 | [25] |
| | HH-H3 | [25] |
| | ROI-H1 | [25] |
| | ROI-H2 | [25] |
| | ROI-H3 | [25] |
| | QRS segment duration | [40] |
| | P-R waves interval | [40] |
| | Q-T waves interval | [40] |
| | T wave interval | [40] |
| | P wave interval | [40] |
| | Weighted centre of the bispectrum (ROI-WCOB) | [35] |
| | Very Low-Frequency band power (VLF) | [23] |

| | | |
|---|---|---|
| Time | Standard Deviation of Average of all NN interval for all 5-minute periods of the entire recording (SDANN) | [36] |
| | ST level | [34] |
| | Standard Deviation of RR intervals (SDRR) | [25], [22], [37], [23] |
| | Mean of RR intervals | [22], [37], [23] |
| | Skewness of RR intervals | [22], [37] |
| | Kurtosis of RR intervals | [22], [37] |
| | Number of adjacent RR intervals differing by more than 50 ms (NN50) | [35] |
| | Sum of NN50 divided by the total number of all RR intervals (PNN50) | [35], [23] |
| | Square root of the mean of the squares of differences between adjacent RR intervals (RMSSD) | [23] |
| | Standard deviation of differences between adjacent RR intervals (SDSD) | [23] |
| | Smoothed Pseudo Winger Ville distribution (SPWVD) | [23] |
| Space | Amplitude of P wave | [34] |
| | Amplitude of Q wave | [34] |
| | Amplitude of R wave | [34] |
| | Amplitude of S wave | [34] |
| | Amplitude of T wave | [34] |
| Nonlinear | Standard Deviation 1 (SD1) | [24], [23] |
| | Standard Deviation 2 (SD2) | [24], [35], [23] |
| | SD1/SD2 ratio | [24], [23] |
| | Sample Entropy | [24], [35] |
| | Approximate Entropy | [22] |

Table 12 shows the horizon of the prediction made by every one of the selected studies, this is, in how much time can the resultant models predict AF episodes.

*Table 12 - Prediction horizon on each one of the selected articles. NR=Not Reported.*

| **Year** | **Study** | **Prediction horizon** |
|---|---|---|
| 2012 | Mohebbi *et al.* [24] | NR |
| 2013 | Costin *et al.* [36] | 30 minutes |
| 2016 | Kim *et al.*[38] | NR |
| 2016 | Shen *et al.*[34] | NR |
| 2016 | Boon *et al.*[25] | 30 minutes |
| 2017 | ElMoaqet *et al.*[22] | 14 days |
| 2018 | Rajalakshmi *et al.*[40] | NR |
| 2018 | Li *et al.*[41] | 2 minutes |
| 2018 | Boon *et al.*[35] | NR |
| 2018 | Ebrahimzadeh *et al.* [23] | 5 minutes |
| 2019 | Attia *et al.*[39] | NR |
| 2019 | Mohamed *et al.*[37] | 60 minutes |

## 4. DISCUSSION

This systematic literature review aims to identify, assess and analyse the recent state-of-the-art of ECG-based models for AF Prediction using Artificial Intelligence techniques. The following paragraphs discuss the previously defined research questions.

### 4.1. How is the prediction problem addressed? (RQ1)

From the selected articles, most of them only address the problem of predicting AF, this is, their main focus is to predict AF and no other types of arrhythmia or heart pathologies.
All the selected articles performed classification prediction, this is, all did classify the prediction with discrete labels.
From all the twelve selected studies, only one performed a risk-based approach on the prediction problem, this is, all the other eleven did a time series prediction of AF. Regarding the number of classes used for the prediction process, only two articles reported a study using a multi-class approach, all the remaining used binary (between "PAF" and "non-PAF" events).
Despite not, all the studies reported the event horizon used for the prediction used method, two of them used a 30 minutes horizon, and the remaining used 14 days, 60 minutes, 5 minutes, 2 minutes and under a 0-minute horizon, this is, immediately before the AF event, each.
Eight out of twelve of the selected studies performed prediction of AF with input signals shorter or equal to 300 seconds, this is, five minutes long, which was also the most used length of signal in all the studies.
When looking at the datasets used by the selected articles, we can see that the three most accurate models are from three of the five studies that used the dataset [32], thus identifying this as a good option for further work on assessing the problem.

### 4.2. What databases and features are used? (RQ2)

Despite some of the selected studies do not perform ECG signal features selection either extraction, when performing it, the selected features directly impact the model's capability of predicting AF existence with higher accuracy.
Table 11 indicates the different features selected by the articles considered in this systematic literature review.
The most used features are Standard Deviation of RR Intervals, Low-frequency band power, Mean of RR Intervals and Standard Deviation, being used by, at least, 3 different selected articles.
Most of the approaches are based solely on ECG signals, but one study combined ECG signal's data with heart morphology data. Almost half of the selected articles used the Atrial Fibrillation Prediction Database ([32]), a quarter of them used a specifically collected dataset of ECG signals. Others used a UCI Repository Warehouse ([29]) dataset, the Mayo Clinic ([30]) ECG Laboratory's database, the Medical Information Mart for Intensive Care III database ([31]), and a China Kadoorie Biobank's ([33]) database.
According to the article [40] the UCI Repository Warehouse dataset used consists of 452 instances with 279 attributes, where the ECG reports are in image format.

The Mayo Clinic ECG Laboratory's database as used by [39], included "all patients aged 18 years or older with at least one digital, normal sinus rhythm, standard 10-second, 12-lead ECG acquired in the supine position" between 1993 and 2017. The signals were acquired at a sampling rate of 500 Hz using a GE-Marquette ECG Machine[30] and stored using the MUSE data management system[31]. All the records were "over-read by a physician-supervised, trained technician, with corrections made to the diagnostic labels as needed".
As used by [37], the Medical Information Mart for Intensive Care III database was collected from 2001 to 2012. It contains information from over 40 thousand patients, about Heart Rate, Arterial Blood Pressure and Respiration. This database also contains "charts at a higher frequency like ECG and continuous blood pressure from Intensive Care Units patients". For the study, only the patients who have developed AF during their recordings are considered.
The studies [23]–[25], [35], [36] used the Atrial Fibrillation Prediction Database, which "consists of excerpts of two-channel long-term ECG (Holter) recordings and is divided into a learning set and a test set of equal size. The database includes the digitized ECG signals (sampled at 128 Hz per signal, with 12-bit resolution) and a set of unaudited, automatically-generated QRS annotations", as in [32]. The records were collected from 48 individuals, although the selected articles always refer to 53 or 75 participants, as in Table 7.
Finally, the study [34] was based on a database from the China Kadoorie Biobank, which is a cohort study of over 520000 adults from 10 different areas from China, collected from 2004 to 2008 using questionnaires and anthropometric and physiological measurements as well as blood samples of every participant. For the study, the 12-lead ECG data of 10 seconds duration at 500 Hz were used, which were collected from 24369 participants using a Mortara ELIx50 device during 2013 and 2014, as well as the blood pressure data (systolic and diastolic).

### 4.3. What pre-processing algorithms are used? (RQ3)

The pre-processing methods used in all the twelve selected studies are presented in Table 9.
Although not all the articles indicate the pre-processing methods applied, due to some of them were elaborated applying prediction methods that do not need any pre-processing of the signal, the most used pre-processing technique is Noise Removal (5 studies), followed by QRS Detection (4 studies) and Correction of Signal (2 studies). Both Normalisation and Missing Value Removal methods were applied by one study each.

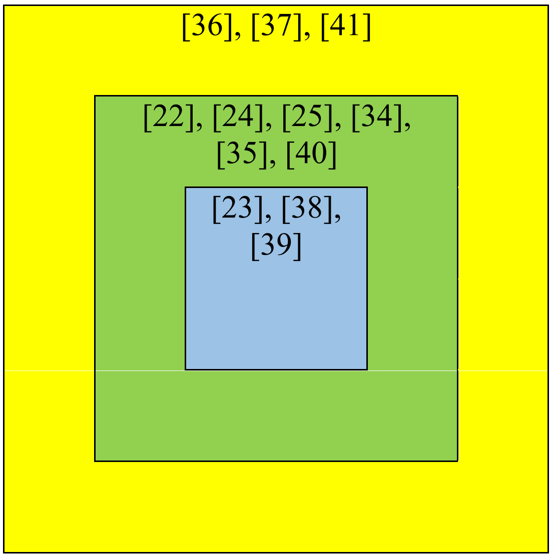
### 4.4. What predictive algorithms are used? (RQ4)

The most used prediction method/algorithm is Support Vector Machine ([22], [24], [25], [34], [35]), followed by Convolutional Neural Network ([38], [39]).
Some other selected studies applied either statistical AI methods (HRV analysis and Morphologic Variability of QRS complexes, Belief Functions Theory), or Arrhythmia Fuzzy Hybrid Classifier, Markov Chain, or, at last, Mixture of Experts.

[30] "MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020]
[31] MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020]

Dividing the predictive algorithms into three classes, this is, Deep Learning, Machine Learning, and Artificial Intelligence, we can identify the type of prediction approach executed by each one of the selected studies, as presented in Table 13.

*Table 13 - Class of Artificial Intelligence methods applied by the selected studies. Yellow=Artificial Intelligence, Green=Machine Learning, Blue=Deep Learning.*




### 4.5. Which are the best models that perform the best? (RQ5)

To address a comparative evaluation of the models used by the selected studies, the authors of this systematic literature review cluster the discussion in terms of:

1. Studies using the same datasets;
2. Studies applying the same prediction method or algorithm;
3. Studies based on the same input signal duration;
4. Studies within the same class of Artificial Intelligence applied method (according to Table 13);
5. All the studies.

1. From all the selected studies, only five of them used the same dataset, leaving all the remaining ones working with a dataset that only they used.
Thus, and comparing all the studies that used the Atrial Fibrillation Prediction Database ([32]), three of them achieved accuracies above or equal to 90% by applying (ordered by accuracy level decreasing) Mixture of Experts, Support Vector Machine, and Statistical AI methods ([23], [24], [36]). The two worst performing studies both used Support Vector Machine ([25], [35]), thus not being possible to indicate what was the best method to apply.
2. Regarding studies applying Support Vector Machine, those who perform the best both used as features LF, SD2 and Sample Entropy ([24], [35]).
When looking at the studies that applied Convolutional Neural Networks as a prediction method, both acquired very similar accuracy rates ([38], [39]).
3. Relatively to the studies based on the input of signals with 300 seconds length, the authors of this systematic literature review highlight the article that used as method Mixture of Experts ([23]), also linking the two best performances with the usage of the features LF, HF, SD1, SD2, SD1/SD2 and Sample Entropy ([23], [24]).
In the studies using signals of 30 seconds ([22], [38]), both performed around 85% of accuracy, but the second achieved higher performance, having a higher amount of

individuals from whom the data was collected, as well as applying Support Vector Machine instead of Convolutional Neural Network as the first.
From the two articles that report work done with signals with 10 seconds length ([34], [39]), the second performed better than the first, and applied Convolutional Neural Network method instead of Support Vector Machine, as well as having a higher number of individuals from whom the data was collected (approximately 5,25 times).
4. From the two studies that applied Deep Learning methods ([38], [39]), both acquired very similar accuracy rates.
Looking into the articles working with Machine Learning methods (excluding those who apply Deep Learning techniques) ([22]–[25], [34], [35], [40]), the two that outperformed all the others, achieving accuracies above 95%, used the Atrial Fibrillation Prediction Database, worked with signals of 300 seconds long and with frequency-domain, time-domain and non-linear features extracted from the input ECG signals.
5. The results of the studies revealed that the increase in the length of the period of ECG signal sent for prediction does not necessarily increase the accuracy of the model created. The best prediction accuracies were obtained in the studies [23] (98,21%), [24] (96,64%) and [36] (90,00%), in which there were used signal parts of 300 seconds. Contrasting, the worst accuracies achieved by the models from the selected articles were obtained on the studies [37] (70,49%), [34] (75,60%) and [25] (80,20%), with signal durations of 3600, 10 and 1800 seconds respectively. These data can indicate that signals too short (10 seconds only) or too long (1800 seconds or above) are not the best approach to the problem being assessed in this systematic literature review.
At last, from the results from the three studies that applied Artificial Intelligence methods ([36], [37], [41]), the authors highlight the achieved accuracy of the first study, which worked with Atrial Fibrillation Prediction Database, having signals with 300 seconds long instead of 120 (second study) or 3600 (third study), performing better among the three.
At last, the authors highlight the achieved accuracy of the study [36], that worked with Atrial Fibrillation Prediction Database, with ECG signals 300 seconds long instead of 120 (as used on the study [37]), or 3600 (on the study [41]), performing. All these three studies used Artificial Intelligence methods.

## 5. CONCLUSION

The present systematic literature review presents and summarizes the current data-based work on predicting Atrial Fibrillation (AF) using Electrocardiogram (ECG) data as input and Artificial Intelligence (AI) methods. Twelve studies were analysed, and the main findings are summarized as follows:

- (RQ1) Despite not existing a current high number of articles published based on studies focused on prediction of AF using AI and ECG signals, most of the existing ones assess the problem by predicting only AF cases, not spending time in the prediction of other cardiovascular issues at the same time, thus being the major number of studies a binary prediction system. The higher part of the existing studies worked with ECG signals 300 seconds long, that is, five minutes. Although some studies tried increasing the length of the period of ECG signal used as input for the prediction models, it does not necessarily increase the accuracy of the obtained final model;

- (RQ2) From all the studies selected for this systematic literature review, the most accurate models were achieved using the Atrial Fibrillation Prediction Database for training. This database was also the most used, by almost half of all the selected articles. The most used features are Standard Deviation of RR Intervals, Low-Frequency band power, Mean of RR Intervals and Standard Deviation, all collected from the ECG signal inputted;
- (RQ3) Among all the selected articles, there were applied many pre-processing techniques, being the most used the Noise Removal, followed by the QRS complex detection;
- (RQ4) The trend in predictive methods based on Machine Learning techniques is increasing. From all the selected studies, the two most used methods were Support Vector Machine and Convolutional Neural Network, being the trend the Machine Learning techniques. However, the authors of this systematic literature review noticed that the usage of deep learning techniques is yet not highly accurate when comparing to simpler Support Vector Machine methods;
- (RQ5) Generally, the models based on Machine Learning methods achieved higher accuracy rates. The higher accuracy was obtained by applying a Mixture of Experts method, followed by s Support Vector Machine implementation. The selected features that conducted to higher accuracy were LF, SD2 and Sample Entropy. Also, the usage of ECG signals 300 seconds long as input for the method's training led to a high rate of prediction accuracy. The database that conducted all the three most accurate models achieved was the Atrial Fibrillation Prediction Database.

As shown by Figure 1, between 2009 and 2019 (the period this systematic literature review covers), more than 80% of the total published studies were performed from 2016 ahead, 50% belonging to the last two years (2018 and 2019).

The amount of work on the prediction of AF episodes is rapidly increasing and showing promising results. Although deep learning methods have already shown outstanding results on the prediction of several areas, namely healthcare, but were not yet applied to many studies, this is, focusing on the prediction of AF using ECG signals. The best results tend to be achieved using Machine Learning and Deep Learning techniques, namely Support Vector Machine and Mixture of Experts.

At last, some limitations of this systematic literature review should be mentioned.

First, this systematic literature review only concerned research in papers written in English. Second, the research for articles returned few articles, even with a cross-reference of the selected studies. Third, this review excluded all the studies that included data collected from patients with recent surgical proceedings or with known cardiovascular conditions that could infer the results of an ECG exam. Finally, the selected studies had to contain evaluation measurements such as accuracy, sensitivity, specificity, or the confusion matrix, excluding any article without any of these evaluations.

**Acknowledgements**

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020 (Este trabalho é financiado pela FCT/MCTES através de fundos nacionais e quando aplicável cofinanciado por fundos comunitários no âmbito do projeto UIDB/EEA/50008/2020).

This article/publication is based on work from COST Action IC1303 - AAPELE - Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226 - SHELD-ON - Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). More information in www.cost.eu.

**Conflict of interest**

The authors declare no conflict of interest.

**REFERENCES**

[1] WHO, "WHO | About cardiovascular diseases," *WHO*, 2011. [Online]. Available: https://www.who.int/cardiovascular_diseases/about_cvd/en/%0Ahttp://www.who.int/mediacentre/factsheets/fs317/en/%0Ahttp://www.who.int/cardiovascular_diseases/about_cvd/en/. [Accessed: 29-Jan-2020].
[2] W. H. Organization, "Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016," Geneva, 2018.
[3] E. Wilkins, W. L., K. Wickramasinghe, and B. P, "European Cardiovascular Disease Statistics 2017 edition," *Eur. Hear. Netw.*, pp. 8–15; 94, 118, 127, 149, 162, 174, 2017.
[4] J. Leal, R. Luengo-Fernández, A. Gray, S. Petersen, and M. Rayner, "Economic burden of cardiovascular diseases in the enlarged European Union," *Eur. Heart J.*, vol. 27, no. 13, pp. 1610–1619, Jul. 2006, doi: 10.1093/eurheartj/ehi733.
[5] D. J. Gladstone *et al.*, "Atrial fibrillation in patients with cryptogenic stroke," *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2467–2477, Jun. 2014, doi: 10.1056/NEJMoa1311376.
[6] T. Sanna *et al.*, "Cryptogenic stroke and underlying atrial fibrillation," *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2478–2486, Jun. 2014, doi: 10.1056/NEJMoa1313600.
[7] Mayo Clinic, "Atrial fibrillation - Symptoms and causes - Mayo Clinic," 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624. [Accessed: 29-Jan-2020].
[8] Mayo Clinic Staff, "Electrocardiogram (ECG or EKG) - Mayo Clinic," *Mayo Clinic*, 2018. [Online]. Available: https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983. [Accessed: 29-Jan-2020].
[9] I. Matias, N. Pombo, and N. Garcia, "Towards a fully automated bracelet for health emergency solution," in *IoTBDS 2018 - Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security*, 2018, vol. 2018-March, pp. 307–314, doi: 10.5220/0006800803070314.
[10] N. Pombo, P. Araújo, J. Viana, and M. D. Da Costa, "Evaluation of a ubiquitous and interoperable computerised system for remote monitoring of ambulatory post-operative pain: A randomised controlled trial," *Technol. Heal. Care*, vol. 22, no. 1, pp. 63–75, Jan. 2014, doi: 10.3233/THC-130774.
[11] N. Pombo and N. M. Garcia, "UbiSleep: An ubiquitous sensor system for sleep monitoring," in *International Conference on Wireless and Mobile Computing, Networking and Communications*, 2016, doi: 10.1109/WiMOB.2016.7763192.
[12] N. Pombo, P. Araújo, J. Viana, B. Junior, and R. Serrano, "Contribution of web services to improve pain diaries experience," in *Lecture Notes in Engineering*

*and Computer Science*, 2012, vol. 2195, pp. 589–592.
[13] H. P. Da Silva, J. Guerreiro, A. Lourenço, A. Fred, and R. Martins, "BITalino: A novel hardware framework for physiological computing," in *PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems*, 2014, pp. 246–253, doi: 10.5220/0004727802460253.
[14] T. Lysaght, H. Y. Lim, V. Xafis, and K. Y. Ngiam, "AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research," *Asian Bioeth. Rev.*, vol. 11, no. 3, pp. 299–314, Sep. 2019, doi: 10.1007/s41649-019-00096-0.
[15] S. Higgins, Julian PT and Green, *Cochrane Handbook for Systematic Reviews of Interventions*. 2019.
[16] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *PLoS Medicine*, vol. 6, no. 7. p. e1000100, 21-Jul-2009, doi: 10.1371/journal.pmed.1000100.
[17] D. Moher *et al.*, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Annals of Internal Medicine*, vol. 151, no. 4. American College of Physicians, pp. 264–269, 2009, doi: 10.7326/0003-4819-151-4-200908180-00135.
[18] F. Weidt and R. Silva, "Systematic Literature Review in Computer Science-A Practical Guide," *Relatórios Técnicos do DCC/UFJF*, vol. 1, no. 0, pp. 1–7, 2016, doi: 10.1027/1016-9040.11.3.244.
[19] "The best research databases for computer science [Update 2019] - Paperpile." [Online]. Available: https://paperpile.com/g/research-databases-computer-science/. [Accessed: 12-Apr-2020].
[20] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan-a web and mobile app for systematic reviews," *Syst. Rev.*, vol. 5, no. 1, p. 210, Dec. 2016, doi: 10.1186/s13643-016-0384-4.
[21] "5: Inclusion and exclusion criteria - Systematic Reviews - Library Guides at University of Melbourne." [Online]. Available: https://unimelb.libguides.com/c.php?g=492361&p=3368110. [Accessed: 15-Jan-2020].
[22] H. ElMoaqet, Z. Almuwaqat, M. Ryalat, and N. Almtireen, "A new algorithm for short term prediction of persistent atrial fibrillation," *2017 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol.*, pp. 1–6, 2017.
[23] E. Ebrahimzadeh *et al.*, "Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal," *Comput. Methods Programs Biomed.*, vol. 165, pp. 53–67, 2018.
[24] M. Mohebbi and H. Ghassemian, "Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal," *Comput. Methods Programs Biomed.*, vol. 105, no. 1, pp. 40–49, 2012.
[25] K. H. Boon, M. Khalil-Hani, M. B. Malarvili, and C. W. Sia, "Paroxysmal atrial fibrillation prediction method with shorter HRV sequences," *Comput. Methods Programs Biomed.*, vol. 134, pp. 187–196, 2016.
[26] R. Spronken-Smith, "SJR : Scientific Journal Rankings," *Assessment & Evaluation in Higher Education*, 2012. [Online]. Available: https://www.scimagojr.com/journalrank.php?country=PT. [Accessed: 18-Jan-2020].

[27] B. Schauerte, "Conference Ranks." 2017.
[28] "Scopus - Document search | Signed in." [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 15-Jan-2020].
[29] A. Frank and A. Asuncion, "{UCI} Machine Learning Repository." 2010.
[30] Mayo Clinic, "Mayo Clinic Mayo Clinic." 2019.
[31] M. Laboratory For Computational Physiology, "The MIMIC III Clinical Database," 2015, doi: 10.13026/C2XW26.
[32] G. B. Moody, A. L. Goldberger, S. McClennen, and S. P. Swiryn, "Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001," *Comput. Cardiol.*, pp. 113–116, 2001, doi: 10.1109/CIC.2001.977604.
[33] R. Questionnaire, "China Kadoorie Biobank," 2013. [Online]. Available: https://www.ckbiobank.org/site/. [Accessed: 18-Jan-2020].
[34] Y. Shen, Y. Yang, S. Parish, Z. Chen, R. Clarke, and D. A. Clifton, "Risk prediction for cardiovascular disease using ECG data in the China kadoorie biobank," *2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 2419–2422, 2016.
[35] K. H. Boon, M. Khalil-Hani, and M. B. Malarvili, "Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III," *Comput. Methods Programs Biomed.*, vol. 153, pp. 171–184, 2018.
[36] H. Costin, C. Rotariu, A. Păsărică, and ru, "Atrial fibrillation onset prediction using variability of ECG signals," *2013 8th Int. Symp. Adv. Top. Electr. Eng.*, pp. 1–4, 2013.
[37] M. Mohamed, M. Farah, and A. Fahed, "Evidence-based prediction of Atrial Fibrillation using physiological signals," *2019 3rd Int. Conf. Bio-engineering Smart Technol.*, pp. 1–4, 2019.
[38] J. Kim, O. Sangjun, Y. Kim, and M. Lee, "Convolutional Neural Network with Biologically Inspired Retinal Structure," *7th Annu. Int. Conf. Biol. Inspired Cogn. Archit. BICA 2016, held July 16 to July 19, 2016 New York City, NY, USA*, vol. 88, pp. 145–154, 2016.
[39] Z. I. Attia *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.
[40] S. Rajalakshmi and K. V Madhav, "A collaborative prediction of presence of Arrhythmia in human heart with electrocardiogram data using machine learning algorithms with analytics," *J. Comput. Sci.*, vol. 15, no. 2, pp. 278–287, 2019.
[41] Z. Li, H. Derksen, J. Gryak, H. Ghanbari, P. Gunaratne, and K. Najarian, "A Novel Atrial Fibrillation Prediction Algorithm Applicable to Recordings from Portable Devices," *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4034–4037, 2018.
[42] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985, doi: 10.1109/TBME.1985.325532.
[43] MATLAB, "MATLAB documentation," *MATLAB*. p. R2012b, 2012, doi: 10.1201/9781420034950.
[44] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 2014, pp. 675–678, doi: 10.1145/2647868.2654889.
[45] C. C. Chang and C. J. Lin, "LIBSVM: A Library for support vector machines,"

*ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, doi: 10.1145/1961189.1961199.

[46] International Organization for Standardization, “International Standard ISO/IEC 14882:2014(E) – Programming Language C++,” p. 1358, 2014.

[47] et al. Abadi, Mart'in, Barham P, Chen J, Chen Z, Davis A, Dean J, “TensorFlow: A system for large-scale machine learning,” in *12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16)*, 2016, pp. 265–83.

## “Prediction of Paroxysmal Atrial Fibrillation onset on Electrocardiogram with Machine Learning and Spectrograms”

Submitted

# Prediction of Paroxysmal Atrial Fibrillation onset on Electrocardiogram with Machine Learning and Spectrograms

Igor Matias[a,b,32], Nuno Garcia[a,b], Sandeep Pirbhulal[a,c], Virginie Felizardo[a,b], Nuno Pombo[a,b], Henriques Zacarias [a,b], Miguel Sousa[a], Eftim Zdravevski[d,e], Ana Gouveia[a]

[a]*Universidade da Beira Interior, Covilhã, Portugal*

[b]*Instituto de Telecomunicações, Covilhã, Portugal*

[c]*Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway*

[d]*Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, Skopje, North Macedonia*

[e]*COPELABS, Universidade Lusófona de Humanidades e Tecnologias, Lisboa, Portugal*

**Abstract:**
**Background:** Atrial Fibrillation is a health condition frequently asymptomatic and thus under-detected but associated with several possible causes of death such as stroke, blood clots, and heart failure. This study aims to develop a new and innovative approach for the prediction of the onset of a Paroxysmal Atrial Fibrillation episode, which is hard to detect and can lead to serious health complications if not correctly diagnosed.
**Methods:** Using signals from the online Physionet Repository open library, we constructed, trained, and tested two deep learning models for the classification of an Electrocardiogram signal of a normal type immediately preceding the onset of a Paroxysmal Atrial Fibrillation episode.
**Results:** The proposed methodology for prediction of a Paroxysmal Atrial Fibrillation onset achieved a maximum accuracy and F1 score of up to 96.73% and 95.06%, respectively.
**Conclusions:** These results are auspicious and can lead to practical applications for timely diagnose of Paroxysmal Atrial Fibrillation episodes.

**Keywords:** ECG waveform, Electrocardiogram, Artificial Intelligence, prediction algorithms, Atrial Fibrillation

## 6. INTRODUCTION

As stated by the World Health Organization (WHO), Cardiovascular Diseases (CVDs) are associated with heart and blood vessels-based disorders. CVDs may include hypertension, heart attack, cerebrovascular disease, heart failure, rheumatic congenital heart disease, and cardiomyopathies, among others [1], which were the most significant cause of death worldwide in 2016 [1], adding to approximately 27.00% of all the

---

[32]Corresponding author.
Email address: igor.matias@ubi.pt.
Postal address: Universidade da Beira Interior, Departamento de Informática, Rua Marquês d'Ávila e Bolama, 6201-001, Covilhã, Portugal.

estimated deaths in the world [2] and about 45.00% in Europe by 2017 [3]. Solely in the European Union, in the year of 2006, CVDs were estimated to have cost €104.00 thousand million for healthcare systems [4].

Among the CVDs, Atrial Fibrillation (AF) is an arrhythmia commonly underdiagnosed, which is characterised by irregular heartbeats, that can lead to blood clots, heart failure, stroke and other heart-related complications including death [5], [6]. It can assume four different types, Paroxysmal AF, Persistent AF, Long-standing Persistent AF and Permanent AF. In [7], A. S. Go *et al.* refer that about 15.00% of strokes occur in people with AF. According to [8], only the severest Long-standing Persistent and Permanent can be easily detected with an Electrocardiogram (ECG) test, the other types being harder to identify due to the irregularity of its symptoms. In a clinical environment, AF presents itself in different forms, often starting as paroxysmal (which is self-terminating) and becoming more persistent over time. Paroxysmal AF (PAF) is defined as attacks of AF, which last from two minutes to less than seven days, spontaneously reverting to normal sinus rhythm.

This irregularity of symptoms makes it very hard to detect the two less severe types of AF, due to the high probability of these patients not presenting any symptoms during an ECG test. Aiming to avoid this low efficiency of detection of AF, predictive models are being developed for the diagnosis of a patient's AF state based on a short ECG signal, avoiding extra-long and intrusive devices and methodologies.

Nowadays, portable devices such as smartwatches, smart fitness bands or portable medical signal collectors may have a crucial role in evolving the way we diagnose several health disorders before they step into a high-risk pathology. This is due to its ease on recording, for example, ECG and pulse signals [9]–[12], being always present on the patient, and therefore being able to collect data from several moments of the day, recording ECG for different activities and emotional states. Some of these devices, despite using a smaller number of ECG leads, sometimes 3 or 2, have been proved to be as efficient as Hospital grade ECG equipment, as tested in [13].

In the last years, several new methods to detect and to predict the existence of the different types of AF were developed. These new approaches require powerful algorithms combined with innovative sensors, applying several different types of artificial intelligence (AI) methods.

There is a multitude of benefits from integrating AI into healthcare, including task automation and analysing patients' big data sets to deliver better healthcare, faster and at a lower cost [14]. The usage of AI into healthcare, and consequently AF detection and prediction, facilitates fast analysis of bigger datasets, easing the workload of healthcare professionals, and making possibly automated and real-time diagnosis, anytime and anywhere.

In this paper, we propose an accurate PAF onset prediction method applying Machine Learning techniques combined with ECG signal spectrograms and noise reduction methods and present early results. Fig. 1 presents a comparison between the common approach and the proposed approach to this prediction problem.

The article is organised as follows. Section 2 presents relevant related work, and section 3 describes the used dataset for the training and evaluation of the algorithm. Section 4 elaborates the construction, designing and applied methods and models, and section 5 shows the obtained results. Finally, we discuss the findings and conclude the paper in Sections 6 and 7, respectively.

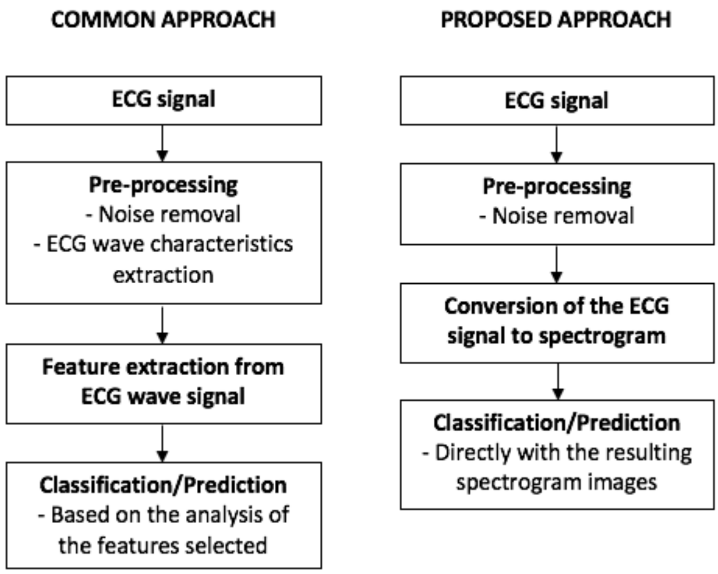



*Figure 24 - Comparison between the common and the proposed approach.*

## 7. RELATED WORK

The detection of AF is commonly performed by analysing the signal collected from an ECG, a non-invasive and painless test with quick results, typically outputting several charts resulting from a 12 lead collection setup [15]. This detection and prediction of the onset of PAF are clinically important due to its increasing possibility of evolving to permanent AF, which makes it impossible the sinus rhythm to be restored to normal functioning. According to [16], about 18.00% of PAF cases evolve to permanent AF over four years. Several prediction methods have been presented in the last years, and many of them are based on the ECG analysis.

The study presented in [17] applied a Support Vector Machine (SVM) technique on the Atrial Fibrillation Prediction Database (AFPD), in an experimental time series approach on the binary classification of ECG signals being pre-PAF onset or normal, achieving an accuracy of 96.64% with an input signal of 300 seconds.

In [18], the authors performed Heart Rate Variability (HRV) analysis and Morphologic Variability of QRS complexes on data from AFPD, by inputting 300 seconds long ECG's into a time series binary classification, obtaining 90.00% accuracy for an event horizon of 30 minutes.

Authors of [19] used its dataset of ECG tests, that was used to train Convolutional Neuronal Network (CNN) with On/Off Rectified Linear Unit (ReLU), obtaining a binary time series classification model with 83.58% of accuracy for 30 seconds input signals.

The authors of [20] developed an SVM approach on the Chine Kadoorie Biobank dataset with a risk-based multi-classification on 10 seconds input signal, correctly classifying between "normal", "ischemia", "arrhythmia", and "hypertrophy" ECG with an accuracy of 75.60%.

Research published in [21] also shows the results of applying SVM techniques on the AFPD dataset, but performing a time series binary classification for an event horizon of 30 minutes, with 1800 seconds input, reporting accuracy of 80.20%.

In [22] is presented the work on applying a Weighted version of SVM on a private dataset of ECG signals from sheep, with an accuracy of 84.90% for a binary time series classification of 30 seconds signals and an event horizon of 14 days.
The authors of the study [23] used an Arrhythmia Fuzzy Hybrid Classifier (ARFC) algorithm on a UCI Repository's dataset for a multi-time series classification approach that has obtained 82.80% of accuracy.
The study published in [24] applied the Markov Chain algorithm for a binary time series classification of collected signals composing a dataset with 120 seconds input signals and a 2 minutes event horizon, reporting accuracy of 82.00%.
Studies [25] and [26] used the AFPD dataset with SVM and Mixture of Experts methods, respectively. The former achieved 87.70% of accuracy on a binary time series classification with an input of 900 seconds, and the latter reached an accuracy of 98.21% on 300 seconds signals for 5 minutes event horizon.
In [27] is presented a binary time series classification approach on data from Mayo Clinic ECG Laboratory using a CNN with 10 seconds signal input, for an obtained accuracy of 83.30%.
The authors of the article [28] achieved a 70.49% accuracy with a method based on the Belief Functions Theory applied to data from the Medical Information Mart for Intensive Care III database, performing a binary time series classification with 3600 seconds input and an event horizon of 60 minutes.
Tabel 1 (in Section 5) provides a summary of the details of the relevant research published in this field.
As a conclusion, we can see the highest accuracy was achieved by [26], while the lowest by [28]. Despite not existing a current high number of articles published based on studies focused on prediction of AF using AI and ECG signals, most of the existing ones assess the problem by predicting only AF cases, not spending time in the prediction of other cardiovascular issues at the same time. From all the related work here presented, the best results were achieved when using machine learning and deep learning techniques, also achieved on the studies using an input length of 300 seconds. The database Atrial Fibrillation Prediction Database (AFPD) [29] was the most used from all the presented work.

## 8. DATASET

This study used the data available at the Atrial Fibrillation Prediction Database (AFPD) [29], publicly available at the Physionet Repository [30]. AFPD consists of ECG signals, each 30 minutes long, from mainly three types: "normal" tests from patients with no diagnosed PAF; "PAF-distant" tests from patients that were PAF diagnosed but did not have a PAF onset during it (at least 45 minutes before and after a PAF-detected episode, when that happened); “pre-PAF" tests from PAF-diagnosed patients where the ECG signal ends precisely before the onset of a PAF episode.
All the data included in this dataset came from ECG tests of 48 different subjects, and each record contains two-channel traces sampled at 128 Hz with a 12-bit sample resolution.
From this database we extracted 50 records of the "normal" type, plus 25 of "PAF-distant", and 25 more of "pre-PAF". In total, we extracted 100 ECG segments, 30 minutes long each. The "normal" samples were used to evaluate our algorithm and results. These extracted records were stored in ".dat" files, which had all the required ECG points.

Each one of the 100 ECG segments has in its data both collected channels, that is, 100 samples contain data for 200 ECG segments with 30 minutes. Thus, the extracted data was demultiplexed, resulting in a total of 200 samples, 100 records of "normal" type, 50 of "PAF-distant" and 50 of "pre-PAF" types.

### 8.1. DATA DEMULTIPLEXING

The original signal files contained in the database were interpreted using the *Python3* [31] library *WFDB* [32], which make possible the extraction of a specific channel from a single data file. As the database datafiles had two channels of signals, we extracted the "channel=[0]" and the "channel=[1]" for every file, thus obtaining both channels separately.

## 9. PROPOSED METHOD

The goal of the presented study is to develop an effective and easy to implement a method for the prediction of the onset of PAF. Fig. 2 shows the block diagram of the proposed method. Step 0 consists of selecting the 30-minute long signal file. Step 1 pre-processes the input ECG signal with smoothing, median filtering and notch filtering. Step 2 performs data segmentation including the slicing of the 30 minutes signal into 30 seconds portions (60 portions per 30 minutes original) and converts each 30 seconds signal into a spectrogram image. Step 3 predicts the PAF onset with a binary CNN classification.
Each one of the described blocks is detailed in the following subsections.

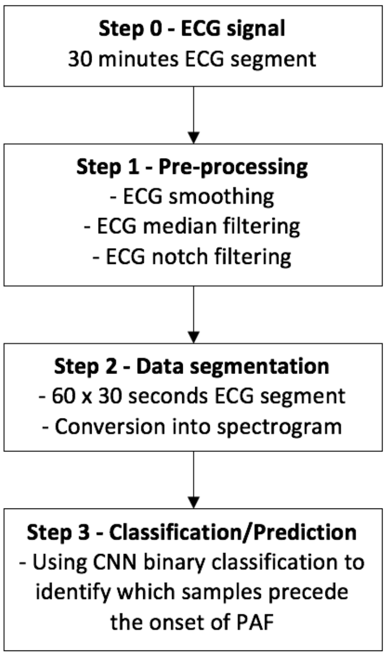



*Figure 25 - Block diagram of the proposed method.*

### 9.1. PRE-PROCESSING

The input signal is a 30 minutes long ECG time series. However, the collected and original data file has some noise and other interferences associated with the signal, making it necessary to process it to obtain a cleaner representation of the signal. The pre-processing stage aims to get the closest to the real data from the ECG signals' datafiles, that is, to remove the maximum possible amount of noise, interference and other phenomenons that could have changed the real and true signal from the body of the patient. Fig. 3 shows the result of each one of the described steps, as applied in [26].

#### 9.1.1. ECG smoothing

The first step of this pre-processing stage is the removal of possible undesired values due to motion of the ECG sensors or the patient's muscle activity, with a simple moving average. We used a window size of 3, according to the formula presented below in (1) and (2), with annotations in (3).
We did a comparison between a window size of 3, 5, 7 and 9, and decided to implement a size of 3 units, because it was the value that maintained the bigger amount of detail of the signal, preventing the changing of the shape of the signal.

$$sum\ of\ neighbor\ points\ (i,N) = \sum_{i-N}^{i+N} K(x) \tag{1}$$

$$smoothed\ signal\ point\ (i) = \frac{1}{2N+1}\ sum\ of\ neighbor\ points\ (i,N) \tag{2}$$

where K is the list of original ECG signal points and N is the parameter of the smoothing function that defines half of the size of the window of observation minus 1, that is

$$N = \frac{length\ of\ the\ observation\ window}{2} - 1. \tag{3}$$

#### 9.1.2. ECG median filtering

Secondly, we applied a median filter for further reduction of the excessive noise of the original signal, with a window size of 3. The median filter is a non-linear digital filtering technique, which replaces each value with the median of the neighbouring values. Equation (4) shown the formula of this filter.
For the window option, we performed a test between the size of 3 and 5 units, in which the lowest value did maintain the highest level of detail of the signal.

$$median\ filtered\ point\ (i,N) = median\ (i-N,\dots,i,\dots,i+N). \tag{4}$$

#### 9.1.3. ECG notch filtering

The third and last step of the pre-processing phase was applying a notch filter to remove the powerline interference from the ECG signal. For that, we used several functions from the *Python3* [31] library *scipy1.4.1* [33], shown in the following equations.

$$(b,a) = iirnotch\left(\frac{cutoff}{\frac{fs}{2}}, q\right) \tag{5}$$

$$notch\ filtered\ signal = lfilter\ (b, a, original\ signal) \quad 6$$

where *iirnotch* and *lfilter* are functions from the library *scipy*, *b* and *a* values are the output of the equation (5), *cutoff* is the frequency we need to remove from the signal, *fs* is the frequency of sampling of the data we use, *q* is the order parameter of the notch filter and *original signal* is the list containing all the points of the original ECG signal to be processed and where the filter is applied.
As stated by [34], a notch filter is widely used in control systems, and, instead of a low-pass filter that attenuates all signals above a certain frequency, a notch filter only removes a narrow band of frequencies, and is also used for removing a signal resonance. In this study, we applied this filter for removal of the frequency of 60 Hz, with the sampling frequency of the used dataset, that is, 128 Hz, and a $2^{\text{-nd}}$ order parameter.
Fig. 3 shows the different steps performed in this phase on an exemplary signal.

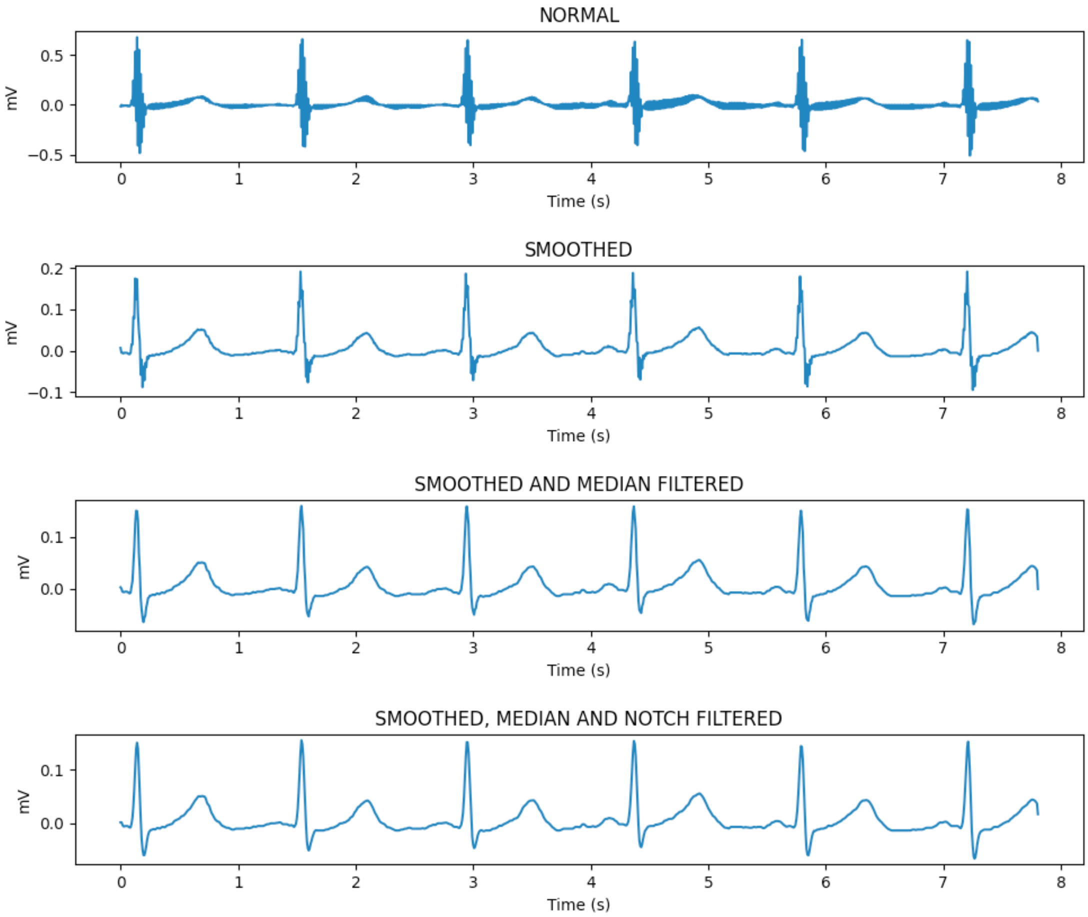



*Figure 26 – Result of each step of the pre-processing of the first seconds of the record "n01" from the database.*

### 9.2. DATA SEGMENTATION

This stage was used on the preparation of the data to input to the Convolutional Neural Network (CNN) model trained in the following step, which comprised the signal cutting in shorter intervals and its conversion into spectrogram images.

### 9.2.1. 60 x 30 seconds ECG segment

As presented at the start of this section, our proposed model used a CNN model training algorithm with spectrogram images as input. However, the data provided in the database were not enough for training a CNN. Thus we had to increase the number of training samples while maintaining the same detail, necessary for a correct and precise prediction of the onset of PAF.
For that, we sliced the original signals, which were 30 minutes long with 128 Hz of sampling frequency, into new ones with only 30 seconds long of the same 128 Hz sampling frequency. At the end of this step, for each one of the 30 minutes signal we had 60 new 30 seconds signals, which scaled our data by 60 times, that is, 6000 samples of "normal" type, 3000 of "PAF-distant" and 3000 more of "pre-PAF".
The Fig. 4 shows the drawing of the first 30 seconds of the database record "p01", which were cut from the original 30 minutes long and stored separately.

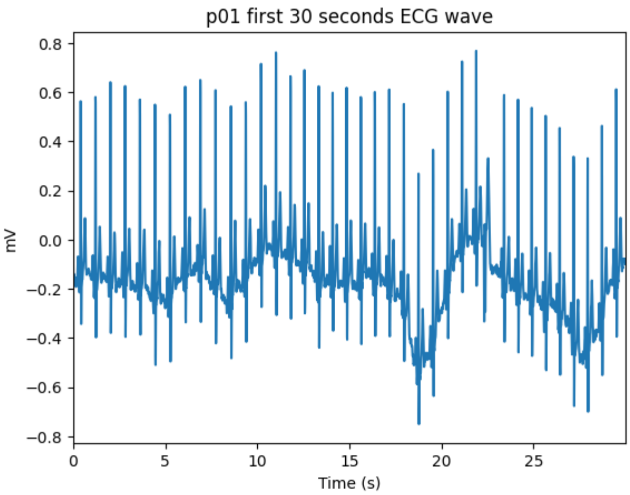



*Figure 27 - Wave signal of the first 30 seconds of the record "p01" from the database.*

### 9.2.2. Conversion into spectrogram

After slicing off the database signals into 30 seconds length samples, we generated power spectrogram images for all signal portions, without normalizing the data.
First, the spectrograms of data were generated using the *Python3* [31] function *specgram* from the library *matplotlib* [35], with the input being the ECG file data and the sampling frequency as 128 Hz. The remaining parameters were not changed, the default values for each one of them being used, including the number of data points used in each block for the *Fast Fourier Transform*, were set to 256.
Fig. 5 shows the produced spectrogram for the wave presented in Fig. 4, that is, the first 30 seconds from the signal "p01" from the database.

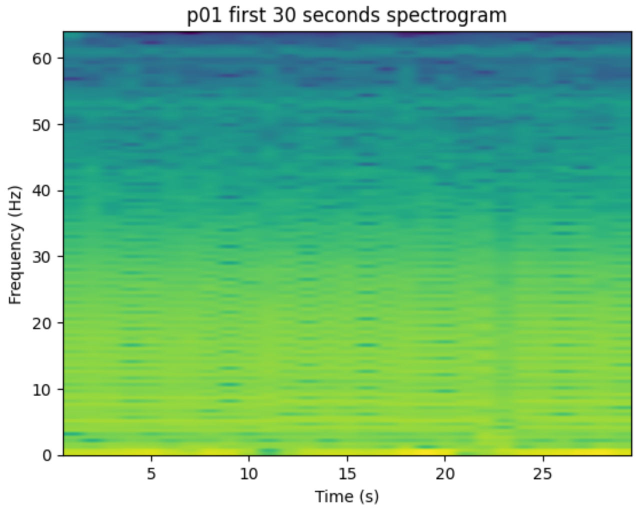



*Figure 28 - Spectrogram of the signal of the first 30 seconds of the record "p01" from the database.*

In it, we can have a better understanding of the signal we need to analyse, by plotting the power of the signal depending on the frequency at each instance of time. If we analyse the spectrogram image as a 3D dimension, where we have time in the X-axis, frequency of the signal in the Y-axis, then the colours of each point in the graph (Z-axis) will indicate us the amount of power the signal has in that specific frequency at a specific time, measured in decibels (dB). The colour bar used in this approach represented the lowest power with a dark blue, and the highest with a bright yellow colour.

This way, we can assure that the first 30 seconds of the record "p01" of the database have a predominance of signals with low frequency, that is, we can easily see that the majority of the yellow regions on the spectrogram are below the frequency of 10 Hz in the 30 seconds signal length.

However, in Fig. 6 and Fig. 7, we can compare the spectrogram images of the last 30 seconds of the records "n01" and "p01", respectively, that is, the last seconds before a normal condition ECG and a PAF episode. It is easily noticeable the difference between the yellow regions between both images, with the majority of signal power on low frequencies for a pre-PAF episode and medium-high frequencies for a normal ECG signal. Figures 8 and 9 also follow this trend, representing the first 30 seconds of the records "n01" and "p01", respectively.

All the figures represent the data as it is at the end of stage 2, that is, ready to be fed to the Convolutional Neural Network described at stage 3. For this reason, there are no axis or white spaces on them, only the spectrogram graph.

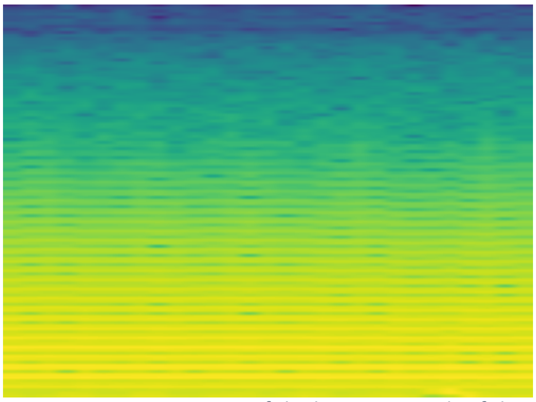


*Figure 29 - Spectrogram of the last 30 seconds of the record "n01".*

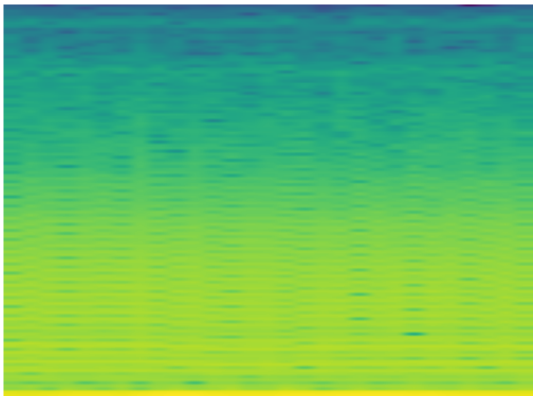


*Figure 30 - Spectrogram of the last 30 seconds of the record "p01".*

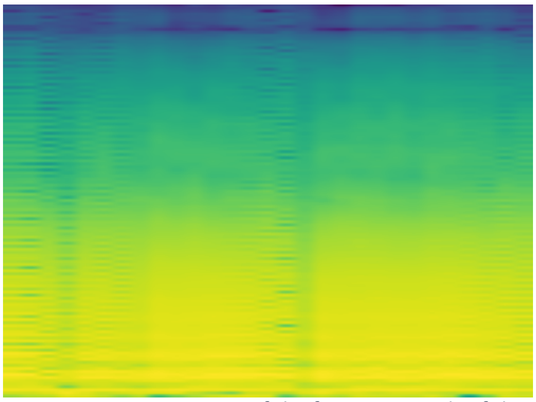


*Figure 31 - Spectrogram of the first 30 seconds of the record "n01".*

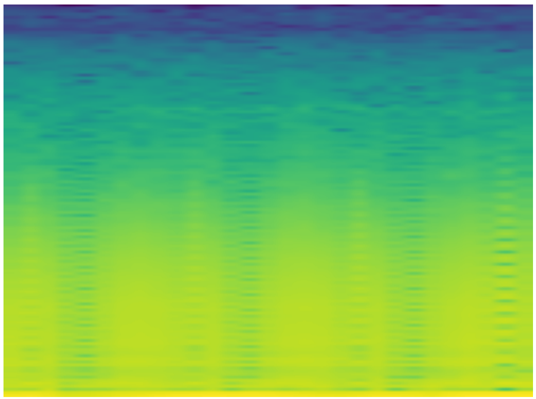


*Figure 32 - Spectrogram of the first 30 seconds of the record "p01".*

### 9.3. CLASSIFICATION/PREDICTION

The next step was constructing, training and testing the CNN model for the classification of the input data. For the prediction of the onset of a PAF episode, our model was trained to classify between samples that preceded or not a PAF event, that is, learning the differences and patterns in data preceding a PAF episode (“pre-PAF” type) and data not preceding, in a short time interval, a PAF episode ("normal" type).

Despite the slicing of the data coming from the used dataset, the resulting amount of data from "pre-PAF" and "normal" types was not enough for a high accuracy as expected initially, which led us onto two different approaches. The first, referred to as the "simple approach", was conducted by training and validating the model with only "pre-PAF" and "normal" data, and testing also with only these two types. In the second, "hybrid approach", we trained and validated the model with all data types, that is, "pre-PAF", "PAF-distant" and "normal" data, and tested with only "normal" and "pre-PAF" data. We calculated the final metrics by evaluating the resulting model only with the data we wanted to distinguish but taking into account the higher amount of data when combining all the three types, which allowed us to better train the algorithm. Both techniques are detailed as follows and were implemented using a Convolutional Neural Network (CNN) with the *Keras Framework* [36] and *Python3* [31].

The data contained in the testing phase was never presented to the model during the training or validation phases, that is, the model only accessed this new data only for testing. This guaranteed the model is robust and can generalise well. In the separation of the training and testing data sets, this was performed using an automated and random

selection procedure coded with *Python3* [31] programming language, thus assuring the models had access to test data from any possible 30 seconds segment of the original ECG signals.
In both approaches, all the input data had the same image size of 496 by 369, and both were trained and tested 10 times with a batch size of 130 and 100 fixed epochs. The obtained results are described in Section 5. A diagram of the network used in both approaches is presented by Fig. 12 and Fig. 13, for the simple and the hybrid approach, respectively.

### 9.3.1. SIMPLE APPROACH

This approach was conducted with three layers of convolution, with 32, 32 and 64 filters by this specific order, and a kernel size of (2,2), ReLU as the activation function and with max-pooling of (2,2) at the end of each one of the layers. This step is known as the Feature Extraction phase of the CNN model, during which the algorithm selects the best features that allow distinguishing, in this specific case, between a "normal" and a "pre-PAF" ECG signal's spectrogram image.
These layers were followed by a flatten instruction, which converts the matrix resulting from the convolution layers into a single array with all the data from these.
In the Classification phase, a fully connected layer with 64 neurons was added with a ReLU function as activation, followed by a dropout of 0.5. The fully connected, or dense, layer was responsible for measuring the importance of each one of the selected features in the classification process, thus adjusting the algorithm to the training data, and the dropout prevented the model from overfitting the training samples, thus generalising better when classifying data different from the training one.
At the end of the model architecture, we used a fully connected layer with only one neuron, and applied a Sigmoid function as activation, thus returning a probability of a spectrogram being “pre-PAF” type or not, so that values above 0.5 represent a binary true classification for ECG signals preceding a PAF episode in the next 29 minutes and 30 seconds or less. The proposed model diagram is represented in Fig. 10.
For compiling this training model, we used the Binary Crossentropy function as loss function and *rmsprop* as optimiser parameter. On the train and test data sets, we applied a rescale of 1/255, and a shear range of 0.2, a zoom range of 0.3 and a true horizontal flip for the training data.
This algorithm was fed with 6704 data samples for training (4454 of “normal” type and 2250 of “pre-PAF” type) and 2235 samples for testing (1485 “normal” and 750 “pre-PAF” type), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing.

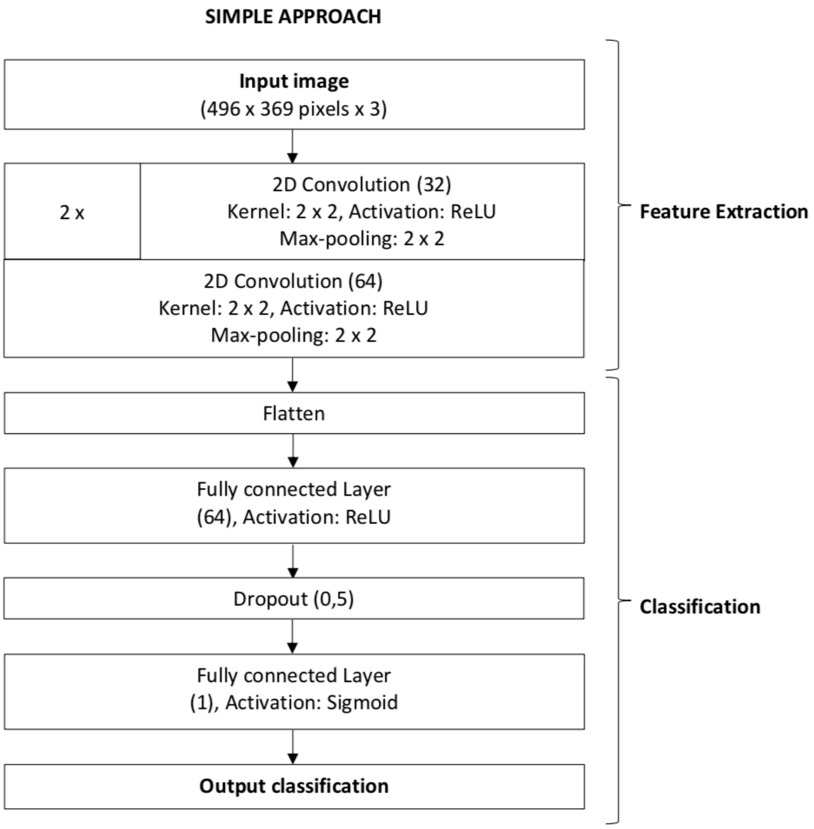



*Figure 33 – Block Diagram of the simple approach model.*

### 9.3.2. HYBRID APPROACH

This hybrid approach was implemented with almost the same specifications as the previous method, except for the number of samples used and the model architecture.
In this approach, we tried to improve the results achieved by training and testing only with "normal" and "pre-PAF" data types. We pre-trained the model with more data ("normal" type as the negative class for PAF prediction, "pre-PAF" and "PAF-distant" types as positive) and tested with only "normal" and "pre-PAF" data, thus facing the low amount of training data issue and validating the model with the same data as on the simple approach. This allowed us to classify an ECG section spectrogram between positive and negative for a preceding PAF episode in the next 29 minutes and 30 seconds, or less.
This approach involved the training and testing of two algorithms, the first one that was trained and tested with all the available data ("normal" as negative, "pre-PAF" and "PAF-distant" data as positive), which resulted in a final model that was used for the training of the second, which was trained and testes with only "normal" and "pre-PAF" data.
On the construction of both algorithms, we designed the two models as having three layers of convolution, with 32, 32 and 64 filters by this specific order, and a kernel size of (2,2), ReLU as the activation function and with max-pooling of (2,2) at the end of each one of the layers on the Feature Selection phase. Next, we coded a flatten instruction (as on the simple approach) and six fully connected layers with 64 neurons with a ReLU function as activation, followed by a dropout of 0.5.

As the final step, as on the simple approach, we implemented a fully connected layer with only one neuron, and applied a Sigmoid function as activation, thus returning a probability of a spectrogram being “pre-PAF” type or not, assuming a true case when the probability goes above 0.5. The proposed model diagram is represented in Fig. 11.
The first algorithm was fed with 8954 data samples for training (4454 of “normal” type and 4500 of “pre-PAF” and “PAF-distant” types combined) and 2985 samples for testing (1485 “normal” and 1500 of “pre-PAF” and “PAF-distant” types combined), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing. This first model was trained with the same specifications as simple and hybrid approach, described above, and the best metrics achieved were chosen as the pre-trained model to be used on the second algorithm of this approach.
The second and main algorithm was fed with 6704 data samples for training (4454 of “normal” type and 2250 of “pre-PAF” type) and 2235 samples for testing (1485 “normal” and 750 “pre-PAF” type), that is, with 75.00% of the total data samples for the training phase and the remaining 25.00% for testing.

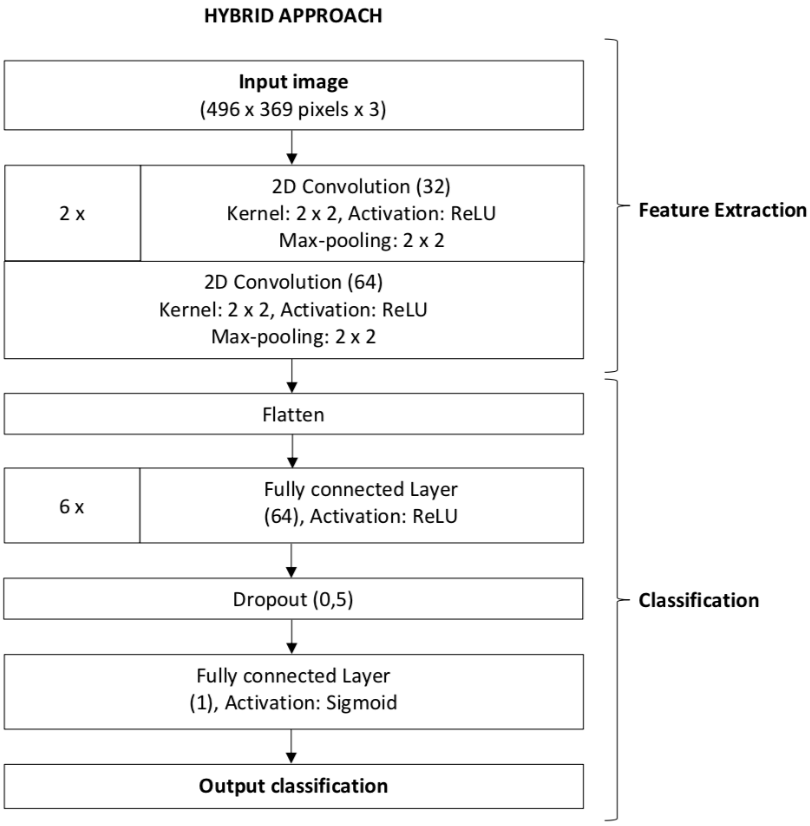



*Figure 34 – Block Diagram of the hybrid approach model.*

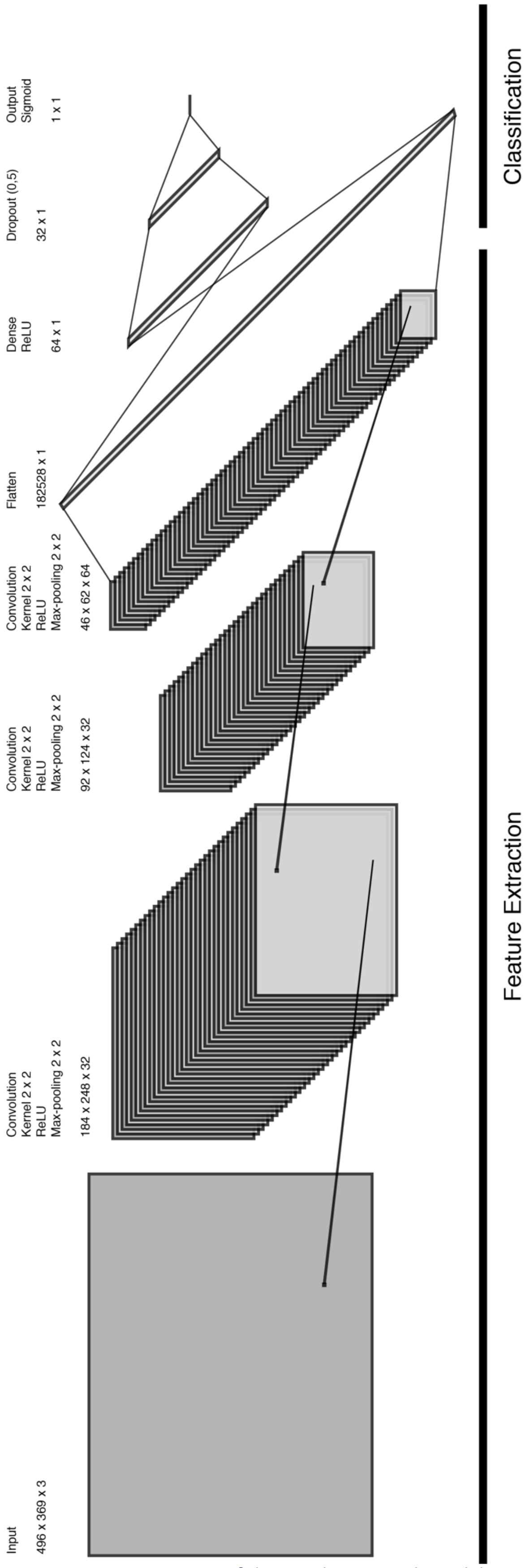



*Figure 35 - Diagram of the simple approach model.*

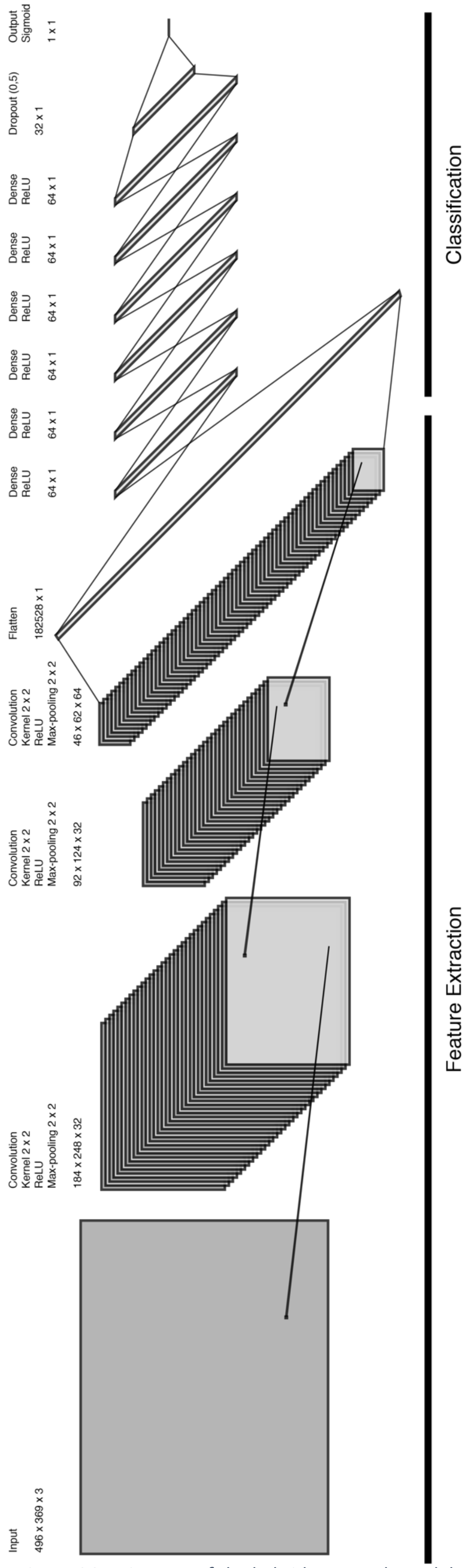



*Figure 36 - Diagram of the hybrid approach model.*

## 10. RESULTS

To evaluate the performance of the proposed methods and obtained classification models, we used the following four measures:

$$Accuracy\ (\%) = \frac{TN + TP}{TN + FP + FN + TP} \times 100 \tag{7}$$

$$Sensitivity\ (\%) = \frac{TP}{TP + FN} \times 100 \tag{8}$$

$$Specificity\ (\%) = \frac{TN}{TN + FP} \times 100 \tag{9}$$

$$F1\ score\ (\%) = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \times 100 \tag{10}$$

where TN, TP, FP and FN stand for True Negative, True Positive, False Positive and False Negative, respectively, and

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN}. \tag{12}$$

If for example a “pre-PAF” episode is classified as a “pre-PAF” episode, then it is said that the episode is classified as TP. On the other hand, if a "normal" episode is classified as "normal", then it is a TN case. Any "normal" episode classified as "pre-PAF" by mistake will produce an FP episode, as well as a "pre-PAF" episode classified as "normal" will produce an FN episode.

| **Author** | **Method** | **Accuracy** | **Sensitivity** | **Specificity** | **F1-Score** |
|---|---|---|---|---|---|
| Mohebbi *et al.* (2012) [17] | SVM | 96.64% | 96.30% | 93.10% | - |
| Costin *et al.* (2013) [18] | HRV analysis and Morphologic Variability of QRS complexes | 90.00% | 89.44% | 89.29% | - |
| Kim *et al.* (2016) [19] | CNN with On/Off ReLU | 83.58% | - | - | - |
| Shen *et al.* (2016) [20] | SVM | 75.60% | - | - | - |
| Boon *et al.* (2016) [21] | SVM | 80.20% | 81.10% | 79.30% | - |

| | | | | | |
|---|---|---|---|---|---|
| ElMoaqet *et al.* (2017) [22] | Weighted SVM | 84.90% | 66.70% | 97.00% | - |
| Rajalakshmi *et al.* (2018) [23] | ARFC algorithm | 82.80% | 0.40% | 0.43% | 1.21% |
| Li *et al.* (2018) [24] | Markov Chain | 82.00% | 86.00% | 80.00% | 74.51% |
| Boon *et al.* (2018) [25] | SVM | 87.70% | 86.80% | 88.70% | - |
| Ebrahimzadeh *et al.* (2018) [26] | Mixture of Experts | 98.21% | 100.00% | 96.55% | - |
| Attia *et al.* (2019) [27] | CNN | 83.30% | 82.30% | 83.40% | 45.40% |
| Mohamed *et al.* (2019) [28] | Belief Functions Theory | 70.49% | 77.07% | 63.90% | - |
| **This research (simple approach)** | **CNN** | **85.01%** (81.61% - 88.19%) | **67.29%** (47.47% - 82.80%) | **93.95%** (85.86% - 98.86%) | **74.77%** (63.40% - 80.95%) |
| **This research (hybrid approach)** | **CNN (pre-trained with different data)** | **95.98%** (94.68% - 96.73%) | **91.09%** (86.40% - 93.60%) | **98.45%** (96.84% - 99.26%) | **93.82%** (91.59% - 95.06%) |

*Table 15 - Summary of different methods for PAF onset prediction and their reported results, compared to this study average metrics.*

In Table 1 the results achieved by each one of the related work studies are presented, in comparison to the results obtained in this research. The metrics resulting of this study were evaluated ten times with the same specifications of the model train and test phases, as described on the Methods section, having achieved an average of 95.98% of accuracy, 91.09% of sensitivity, 98.45% of specificity and 93.82% of F1 score for the hybrid approach, and 85.01% of accuracy, 67.29% of sensitivity, 93.95% of specificity and 74.77% of F1 score for the simple approach.
The metrics for this research in Table 1 present the average results out of 10 training and testing sequences, followed by the worst and best-measured metric in the respective type.

## 11. DISCUSSION

With the results of this study, it is possible to conclude that an ECG of 30 seconds is an appropriate length of a data sample to predict the onset of a PAF episode in a patient. We can also conclude that the spectrogram representation of these signals can maintain the required detail for a neural network to detect some patterns and recognise the difference between a "normal" ECG and a record that will lead to a PAF episode in 29 minutes and 30 seconds or less. This can reduce the computational power required to these technological health approaches, compared to a raw and fully detailed ECG signal, which can overfit and reduce the precision of such algorithms.
The resulting models from this work are two examples of how technology can help to diagnose and prevent several health conditions ahead of time. This can be applied, for example, in an Intensive Care Unit (ICU) tool to alert the physicians for possible dangerous situations happening with a patient.

It can also be used to prevent and reduce False ICU alarms, similar to [37].
This study achieved very high accuracy and precision on the prediction of a PAF episode compared to the previous related works. The proposed models were only slightly worse than [26], which is not very detailed and is challenging to replicate.
However, this study has some limitations that should be mentioned. First, the amount of data available and selected for this particular approach was not large enough for deep learning models. However, the applied data augmentation technique of cutting the original signals into shorter segments proved to be effective. Also, the used dataset can contain errors, kindly reported by the authors in [38]. The conditions of the data labelling were not clear enough and can contain errors, such as some "normal" data samples can be a result of a labelling mistake, coming from an undiagnosed PAF patient or having some portions of PAF related features that were ignored or mistaken during the classification by the physician in charge.

## 12. CONCLUSIONS

Two practical and precise deep learning-based algorithms for predicting the onset of a PAF attack were presented in this article. The achieved high accuracy, specificity and sensitivity were obtained by combining some simple, yet effective and reliable noise reduction techniques and innovative use of spectrogram images for the conversion of its resulting ECG exam signals.
One possible way to further improve this research topic would be, at first, on increasing the amount of data available for the models to train, by merely collecting new ones or by using other datasets. Likewise, from the datasets one could cut the ECG portions immediately before the onset of a PAF episode, which could address for one of the most significant limitations of this work, due to the enormous amount of data related to PAF episodes detection, that is, ECG data containing the PAF episode itself and some signal before or after it. Furthermore, the two proposed approaches leave open the possibility of combining other algorithms with these pre-trained models. This could lead to a broader generalisation of the resulting method, which is one of the most significant limitations on using deep learning algorithms, that is, algorithms that easily recognise and detect patterns on the data in which they were trained but find it very hard to generalise for new and unseen data.
The obtained results of this study demonstrate that its methods can be used as an accurate and reliable tool for the prediction of the onset of PAF events.

**Acknowledgements**

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020 (Este trabalho é financiado pela FCT/MCTES através de fundos nacionais e quando aplicável cofinanciado por fundos comunitários no âmbito do projeto UIDB/EEA/50008/2020).
This article/publication is based on work from COST Action IC1303 - AAPELE - Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226 - SHELD-ON - Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). More information in www.cost.eu.
The authors declare no conflict of interest.

## REFERENCES


[1] WHO, "WHO | About cardiovascular diseases," *WHO*, 2011. [Online]. Available: https://www.who.int/cardiovascular_diseases/about_cvd/en/%0Ahttp://www.who.int/mediacentre/factsheets/fs317/en/%0Ahttp://www.who.int/cardiovascular_diseases/about_cvd/en/. [Accessed: 29-Jan-2020].

[2] W. H. Organization, "Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016," Geneva, 2018.

[3] E. Wilkins, W. L., K. Wickramasinghe, and B. P, "European Cardiovascular Disease Statistics 2017 edition," *Eur. Hear. Netw.*, pp. 8–15; 94, 118, 127, 149, 162, 174, 2017.

[4] J. Leal, R. Luengo-Fernández, A. Gray, S. Petersen, and M. Rayner, "Economic burden of cardiovascular diseases in the enlarged European Union," *Eur. Heart J.*, vol. 27, no. 13, pp. 1610–1619, Jul. 2006, doi: 10.1093/eurheartj/ehi733.

[5] D. J. Gladstone *et al.*, "Atrial fibrillation in patients with cryptogenic stroke," *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2467–2477, Jun. 2014, doi: 10.1056/NEJMoa1311376.

[6] T. Sanna *et al.*, "Cryptogenic stroke and underlying atrial fibrillation," *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2478–2486, Jun. 2014, doi: 10.1056/NEJMoa1313600.

[7] A. S. Go *et al.*, "Prevalence of diagnosed atrial fibrillation in adults: National implications for rhythm management and stroke prevention: The anticoagulation and risk factors in atrial fibrillation (ATRIA) study," *J. Am. Med. Assoc.*, vol. 285, no. 18, pp. 2370–2375, May 2001, doi: 10.1001/jama.285.18.2370.

[8] Mayo Clinic, "Atrial fibrillation - Symptoms and causes - Mayo Clinic," 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624. [Accessed: 29-Jan-2020].

[9] I. Matias, N. Pombo, and N. Garcia, "Towards a fully automated bracelet for health emergency solution," in *IoTBDS 2018 - Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security*, 2018, vol. 2018-March, pp. 307–314, doi: 10.5220/0006800803070314.

[10] N. Pombo, P. Araújo, J. Viana, and M. D. Da Costa, "Evaluation of a ubiquitous and interoperable computerised system for remote monitoring of ambulatory post-operative pain: A randomised controlled trial," *Technol. Heal. Care*, vol. 22, no. 1, pp. 63–75, Jan. 2014, doi: 10.3233/THC-130774.

[11] N. Pombo and N. M. Garcia, "UbiSleep: An ubiquitous sensor system for sleep monitoring," in *International Conference on Wireless and Mobile Computing, Networking and Communications*, 2016, doi: 10.1109/WiMOB.2016.7763192.

[12] N. Pombo, P. Araújo, J. Viana, B. Junior, and R. Serrano, "Contribution of web services to improve pain diaries experience," in *Lecture Notes in Engineering and Computer Science*, 2012, vol. 2195, pp. 589–592.

[13] H. P. Da Silva, J. Guerreiro, A. Lourenço, A. Fred, and R. Martins, "BITalino: A novel hardware framework for physiological computing," in *PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems*, 2014, pp. 246–253, doi: 10.5220/0004727802460253.

[14] T. Lysaght, H. Y. Lim, V. Xafis, and K. Y. Ngiam, "AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research," *Asian Bioeth. Rev.*, vol. 11, no. 3, pp. 299–314, Sep. 2019, doi: 10.1007/s41649-019-00096-0.

[15] Mayo Clinic Staff, “Electrocardiogram (ECG or EKG) - Mayo Clinic,” *Mayo Clinic*, 2018. [Online]. Available: https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983. [Accessed: 29-Jan-2020].

[16] S. M. Al-Khatib, W. E. Wilkinson, L. L. Sanders, E. A. McCarthy, and E. L. C. Pritchett, “Observations on the transition from intermittent to permanent atrial fibrillation,” *Am. Heart J.*, vol. 140, no. 1, pp. 142–145, Jul. 2000, doi: 10.1067/mhj.2000.107547.

[17] M. Mohebbi and H. Ghassemian, “Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal,” *Comput. Methods Programs Biomed.*, vol. 105, no. 1, pp. 40–49, 2012.

[18] H. Costin, C. Rotariu, A. Păsărică, and ru, “Atrial fibrillation onset prediction using variability of ECG signals,” *2013 8th Int. Symp. Adv. Top. Electr. Eng.*, pp. 1–4, 2013.

[19] J. Kim, O. Sangjun, Y. Kim, and M. Lee, “Convolutional Neural Network with Biologically Inspired Retinal Structure,” *7th Annu. Int. Conf. Biol. Inspired Cogn. Archit. BICA 2016, held July 16 to July 19, 2016 New York City, NY, USA*, vol. 88, pp. 145–154, 2016.

[20] Y. Shen, Y. Yang, S. Parish, Z. Chen, R. Clarke, and D. A. Clifton, “Risk prediction for cardiovascular disease using ECG data in the China kadoorie biobank,” *2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 2419–2422, 2016.

[21] K. H. Boon, M. Khalil-Hani, M. B. Malarvili, and C. W. Sia, “Paroxysmal atrial fibrillation prediction method with shorter HRV sequences,” *Comput. Methods Programs Biomed.*, vol. 134, pp. 187–196, 2016.

[22] H. ElMoaqet, Z. Almuwaqat, M. Ryalat, and N. Almtireen, “A new algorithm for short term prediction of persistent atrial fibrillation,” *2017 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol.*, pp. 1–6, 2017.

[23] S. Rajalakshmi and K. V Madhav, “A collaborative prediction of presence of Arrhythmia in human heart with electrocardiogram data using machine learning algorithms with analytics,” *J. Comput. Sci.*, vol. 15, no. 2, pp. 278–287, 2019.

[24] Z. Li, H. Derksen, J. Gryak, H. Ghanbari, P. Gunaratne, and K. Najarian, “A Novel Atrial Fibrillation Prediction Algorithm Applicable to Recordings from Portable Devices,” *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4034–4037, 2018.

[25] K. H. Boon, M. Khalil-Hani, and M. B. Malarvili, “Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III,” *Comput. Methods Programs Biomed.*, vol. 153, pp. 171–184, 2018.

[26] E. Ebrahimzadeh *et al.*, “Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal,” *Comput. Methods Programs Biomed.*, vol. 165, pp. 53–67, 2018.

[27] Z. I. Attia *et al.*, “An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction,” *Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.

[28] M. Mohamed, M. Farah, and A. Fahed, “Evidence-based prediction of Atrial Fibrillation using physiological signals,” *2019 3rd Int. Conf. Bio-engineering Smart Technol.*, pp. 1–4, 2019.

[29] G. B. Moody, A. L. Goldberger, S. McClennen, and S. P. Swiryn, “Predicting the

onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001," *Comput. Cardiol.*, pp. 113–116, 2001, doi: 10.1109/CIC.2001.977604.

[30] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.

[31] Python Software Foundation, "Python 3.0 Release | Python.org," 2009. [Online]. Available: https://www.python.org/download/releases/3.0/. [Accessed: 23-May-2020].

[32] "wfdb · PyPI." [Online]. Available: https://pypi.org/project/wfdb/. [Accessed: 08-Jun-2020].

[33] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[34] G. Ellis, "Filters in Control Systems," in *Control System Design Guide*, Elsevier, 2012, pp. 165–183.

[35] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 99–104, May 2007, doi: 10.1109/MCSE.2007.55.

[36] Chollet François, "Keras: The Python Deep Learning library," *Keras.Io*, 2015, doi: 10.1086/316861.

[37] P. Lameski, E. Zdravevski, S. Koceski, A. Kulakov, and V. Trajkovik, "Suppression of Intensive Care Unit False Alarms Based on the Arterial Blood Pressure Signal," *IEEE Access*, vol. 5, pp. 5829–5836, 2017, doi: 10.1109/ACCESS.2017.2690380.

[38] "AF Termination Challenge Database v1.0.0." [Online]. Available: https://physionet.org/content/afpdb/1.0.0/. [Accessed: 08-Jun-2020].